Transcriptional regulation of the platelet-derived growth factor alpha receptor (*PDGFRA*) gene during development

Lisbeth Flores-García

Thesis submitted to the University of London
in part fulfilment of the requirements for the degree of
Doctor of Philosophy

Wolfson Institute for Biomedical Research
and Department of Biology,
University College London.

2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The platelet-derived growth factor receptor (alpha subunit, *pdgfra*) is essential for the development of several types of precursor cells and their differentiated progeny; for example, the cranial neural crest cells (CNCC), mesenchyme derived-cells (MSdC) and oligodendrocyte progenitors (OLPs) in the central nervous system. Previous studies suggested that separate *cis*-acting specific regulatory elements (SREs) are required to drive *pdgfra* expression in different types of progenitor cells during development. As a first step towards understanding the control of *pdgfra* transcriptional regulation I set out to identify and characterize SREs in vivo.

I analyzed 16 human *PDGFRA*-YAC/BAC mice transgenic lines, most of which I generated myself, and compared their *PDGFRA* expression patterns with each other and with previously-generated *PDGFRA* transgenics during development. These new transgenes gave me a better understanding of *PDGFRA* transcriptional regulation and allowed me to further refine the locations of many important SREs of this gene. Most of the SREs controlling expression in CNCC, sclerotome progenitors, MSdC, and first and second branchial arches seem to be located in the 3 kb upstream sequence of the *PDGFRA* gene. In addition, the postnatal OLP (pOLP)-SREs appear to lie in a region up to 63 kb downstream of the coding sequence.

Furthermore, I used bioinformatic analysis to search for evolutionarily conserved sequences (CS) in the non-coding genomic sequence of the human *PDGFRA* gene, and found that the 3'-untranslated region has CSs, some of which might be pOLP-SREs

because they are present in chicken, Xenopus and mammals but not zebrafish, in keeping with the observation that OLPs in all tetrapods examined - but not fish - express *pdgfra* in their OLPs.

Overall, my results suggest a 'temporal colinearity' model of SREs for transcriptional activation of the *pdgfra* gene during development and that its expression might be managed by separate embryonic and postnatal regulatory systems.

Acknowledgments

This PhD degree has been possible thanks to an international PhD scholarship from CONACYT, México, and the unconditional financial support from Rubí, México; thanks for your endless faith on me.

Acknowledgements to Gijs Afink and Monica Nister from Karolinska Institutet for their collaboration in this project, and invited me to Stockholm. Thanks to X. Q. Zhang for providing the 2.2kb-LacZ human *PDGFRA* mice embryos for comparison. Special thanks to Bill Richardson for reading and correcting this thesis. From his lab, thanks to Marisa Howell for multiple YAC DNA purifications and for her smile and support. Special mention goes to Marta García Del Barrio for reading and discussing this thesis constructively, encouraging and making Science more fun. Thanks to Ana Mora for double fluorescent *in situ* hybridizations with the human *PDGFRA* probe vs. m*Olig2*, m*NeuN* and GFAP. Thanks to Francoise Jamen for her collaboration in this project, and to Ulla Dennehy and Palma Iannarelli for the pronuclear microinjections of the human *PDGFRA* BACs and CE#1^{+}H2120-βgal plasmid. Acknowledgments to Huiliang Li, Raquel Taveira-Marques and Matthew Grist for their advice on BAC engineering by homologous recombination, used to test a conserved element. Thanks to Richard Pug and Ulla Dennehy for helping with animal care. Thanks to Damith Jayatilake for technical advice but especially for his smile at the lab at the beginning of my PhD.

Special thanks to my friends, Deneb, Edith, Claudia, Lucy, Mar and Seratna, for your love, support, encouragement, and those lovely trips around Europe that gave me strength.

Mar, thanks for being courageous, lived the great tough adventure, and also kept motivating me through it all. During this international challenge, my family was accidentally forced but loved to get involved, so thank you so much for your unconditional love and help to both of us and also for making me feel very proud. From London, to Marta, Cris, Raquel, Lola, Simon, Natassa and Erik, thanks for your friendship, thousand dinners, walks, jokes and unforgettable experiences in London. Finally, I must thank me for not going completely mad on the other side of the Atlantic :). I wanted professional and personal adventure, I got it indeed.

Science, the way mankind has made it, is a competitive, powerful, low paid and obsessive affair; but in case a scientific discovery, in between all that, is in front of us as a beautiful, logic and simple principle of nature, it is one of the most wonderful moments a human being can experience.

LFG

I, Lisbeth Flores-Garcia, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

A mi Madre,

con amor, agradecimiento y admiración.

Eres muy valiente.

Dedicated to my Mother,

with love, gratitude and admiration.

You are very brave.

Contents

Figures and Tables

Abbreviations

bp	base pair
BCNS	Basal cell nevus syndrome
BCCs	Basal cell carcinomas
BAC	Bacterial Artificial Chromosome
CML	Chronic myeloid leukemia
CNCC	Cranial neural crest cells
CNS	Central nervous system
DNA	Deoxyribonucleic acid
DEPC	diethylpyrocarbonate
DIG	Digoxygenin
ECD	Extracellular domain
kb	kilobase
KO	Knockout
LMP	Low melting point
M	Molar
MBP	Myelin basic protein
MN	Motor neuron
MSdC	Mesechyme derived-cells
MsCD	Mesenchymal cell derivatives
mRNA	messenger ribonucleic acid
OL	Oligodendrocyte
OLP	Oligodendrocyte progenitor
PAC	P1-bacterial artificial chromosome
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDGFRA	Human platelet-derived growth factor alpha receptor gene
PDGFRA	Human platelet-derived growth factor alpha receptor protein
PDGFRB	Human platelet-derived growth factor beta receptor gene
Pdgfra	Mouse platelet-derived growth factor alpha receptor gene
Pdgfra	Mouse platelet-derived growth factor alpha receptor protein

pdgfra	Any vertebrate platelet-derived growth factor alpha receptor gene
pdgfra	Any vertebrate platelet-derived growth factor alpha receptor protein
PDGFA (B,C or D)	Platelet-derived growth factor A, B, C or D protein
PdgfA (B,C or D)	Mouse platelet-derived growth factor A, B, C or D protein
PdgfA (B,C or D)	Mouse platelet-derived growth factor A, B, C or D gene
PFGE	Pulsed-field gel electrophoresis
PNS	Peripheral nervous system
Pth	*Patch* mutant mice
RE	Regulatory element
Shh	Sonic hedgehog
ScP	Sclerotome progenitors
SRE	*cis*-specific regulatory element
TGF	Transforming growth factor
UTR	Un-translated region
VZ	Ventricular zone
YAC	Yeast artificial chromosome

Chapter 1
General introduction

In this chapter, I describe the main stages of vertebrate development, with special reference to the mouse. This is followed by a summary of central nervous system (CNS) patterning, neurogenesis, gliogenesis, and the origin of oligodendrocyte progenitors.

I also review the mechanisms involved in the control of gene expression during development and some examples of transcriptional regulation. I describe the platelet-derived growth factor alpha receptor (PDGFRA), its function, gene structure and expression pattern, as well as its roles in oligodendrocyte development, cancer, spina bifida. I review previous studies of its transcriptional control and explain the main purpose and contributions of my own research.

1.1 General description of vertebrate development

The mouse has a life cycle of 9 weeks, from fertilization to sexually mature adult, which is relatively short for a mammal and is one of the reasons why it has become an important model organism for vertebrate development. Fertilization of the egg (approx. 100μm diameter) takes place internally in the oviduct; meiosis is then completed and embryogenesis begins. Cleavage takes place in the oviduct and only after 4 ½ days does the embryo implant into the uterine wall. Gastrulation takes place over the next few days,

and by 10 days after fertilization all the organs have begun to develop. Mammalian embryos rely on nutrients obtained from the mother via the placenta (Fig. 1.1, A and B).

Early cell divisions in mouse are very slow compared with Xenopus and chick, the first occurring about 24 hours after fertilization and subsequent cleavages at about 12-hour intervals. They produce a solid ball of cells, the morula (Fig. 1.1, C). By the 32-cell stage the morula contains about 10 internal cells and more than 20 outer cells.

A special feature of mammalian development is that the early cleavages give rise to two groups of cells- the trophectoderm and the inner cell mass (ICM). The internal cells of the morula give rise to the ICM and the outer cells to the trophectoderm. The trophectoderm will give rise to extra-embryonic structures such as the placenta, while the embryo proper develops from the ICM. At this stage (3½ days gestation) the embryo is known as a blastocyst (Fig. 1.1 D) (Gilbert and Raunio, 1997;Wolpert et al., 1998;Kalthoff, 2001;Nagy et al., 2003).

The course of early post-implantation development of the mouse embryo from around 4½ days to 8½ days appears more complicated than that of the chick, partly because of the need to produce a larger variety of extra-embryonic membranes, and partly because the epiblast from which the embryo will develop is distinctly cup-shaped in the early stages (Fig. 1.1, E) . This is a peculiarity of mouse and other rodent embryos. The epiblasts of human and rabbit embryos, for example, are flat. In essence, however, the development of the mouse embryo is very similar to that of the chick and other mammals.

By 6.5 days after fertilization, gastrulation begins with the formation of the primitive streak which is the first visible sign of the embryo's future axis. The development of the primitive streak in the mouse is similar to that in the chick; first it elongates towards the future anterior end of the embryo, with a condensation of cells at its anterior end corresponding to Hensen's node. Cells migrating anteriorly through the node will form the notochord, and both the notochord and somites form anterior to the node.

Most animals are triploblasts, with three germ layers: 1) the endoderm, which gives rise to the gut and its derivatives, such as the liver and lungs in vertebrates; 2) the mesoderm, which gives rise to the skeleto-muscular system, connective tissues and other internal organs such as the kidney and heart and 3) the ectoderm, which gives rise to the epidermis and nervous system. The boundaries between the different layers can be fuzzy and there are notable exceptions. The neural crest in vertebrates, for example, is ectodermal in origin but gives rise both to neural tissue and to some skeletal and other elements, which would normally be considered mesodermal in origin (Gilbert and Raunio, 1997;Wolpert et al., 1998;Gammill and Bronner-Fraser, 2003).

At around 8½ days (Fig. 1.1, F), neural folds start to form at the anterior end on the dorsal side of the embryo. In fact, two days after the formation of the notochord and somites, the neural plate ectoderm above them begins to invaginate creating the neural tube, the embryonic precursor of the central nervous system (CNS). This process is called neurulation. In these final stages of gastrulation the embryo also undergoes complex folding, in which the embryonic endoderm-initially on the ventral surface of the embryo-

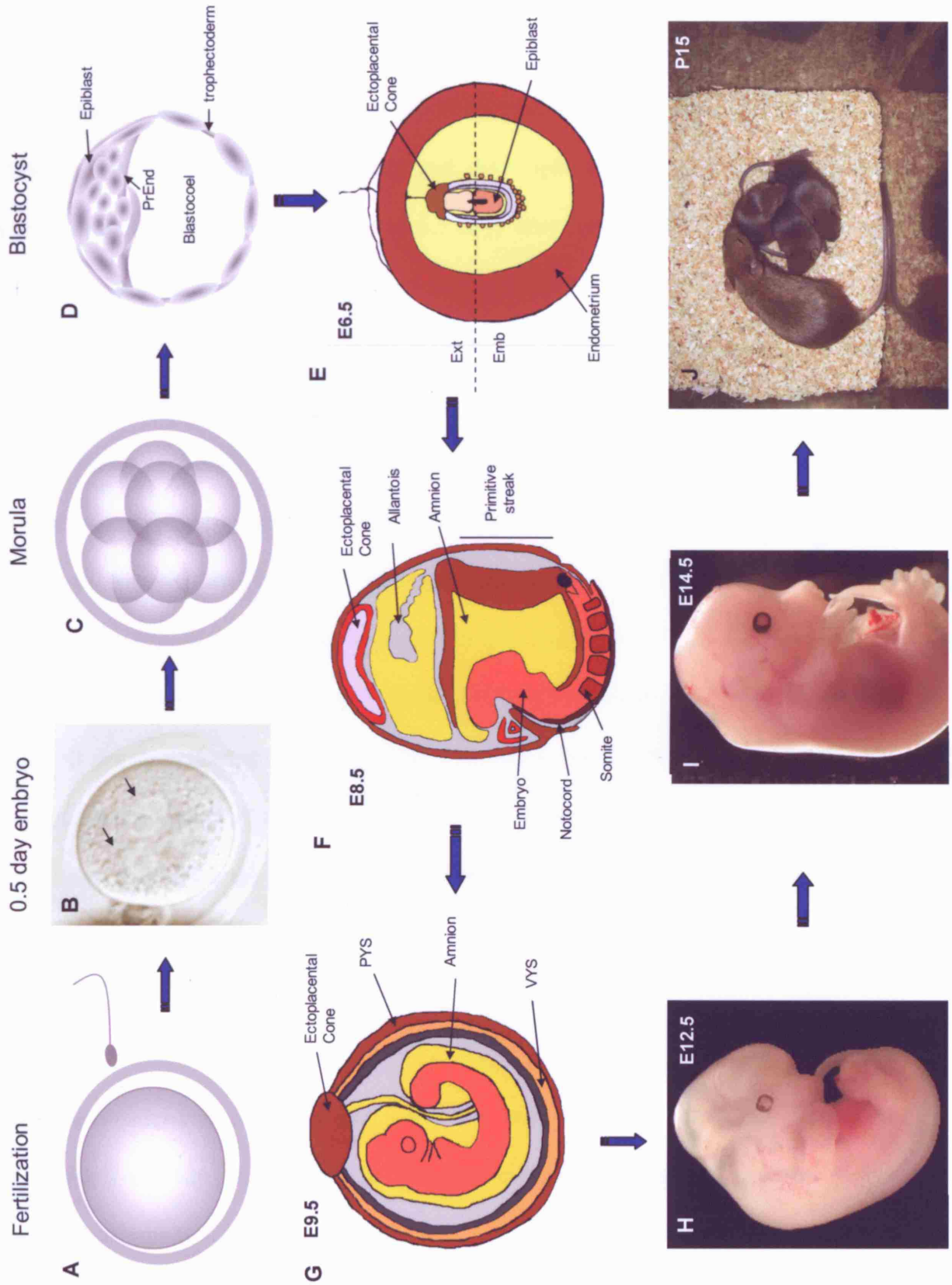

Fertilization
A
0.5 day embryo
B
Morula
C
Blastocyst
D
Epiblast
PrEnd
Blastocoel
trophectoderm
E
E6.5
Ectoplacental Cone
Epiblast
Ext
Emb
Endométrium
F
E8.5
Ectoplacental Cone
Allantois
Amnion
Primitive streak
Embryo
Notocord
Somite
G
E9.5
Ectoplacental Cone
PYS
Amnion
VYS
H
E12.5
I
E14.5
J
P15

Figure 1.1. General representation of the mouse development. The mouse embryo starts after fertilization of the egg by the sperm (A), a 0.5 day embryo have already formed the female and male pronuclei (tiny arrows in B) and also two polar bodies. Morula and blastocyst stages continue (C and D), at the blastocyst stage it is possible to detect the epiblast, the primitive endoderm (PrEnd) and the trophectoderm. Once the blastocyst implants in the female utero (E), at around 6.5 days after fertilization (E6.5), the embryo forms a horizontal axis (dotted line) which shows the cells that will give rise to extra-embryonic (Ext) and embryonic (Emb) tissue. At the E8.5 stage (F), it is finally possible to see an embryo with a head, body and tail shape, structures like notochord and somites appear. However, it is until E9.5 (G) that the embryo finishes rotating to be surrounded by all the extra-embryonic membranes, like the parietal yolk sac (PYS), visceral yolk sac (VYS) and amnion. The embryo continues growing, well protected inside these membranes throughout its development until it is born at E18.5-E19. At E12.5 (H), organs like the heart and limb buds are clearly visible. At E14.5 (I), many organs are developing rapidly and the fingers start to be separated. The pup is born with closed eyelids and no hair but at approximately 15 days postnatal (P15) (J) the pups already have fine hair, small ears and open eyes (photograph of CBA X C57BL/agouti female and three pups).

---

becomes internalized to form the gut, while the heart and liver move into their final positions relative to the gut, and the head becomes distinct. The embryo then turns so that it becomes surrounded by its extra-embryonic membranes. By 9 days, gastrulation is complete: the embryo has a distinct head and the forelimb buds start to develop.

After 9.5 days of embryogenesis (E9.5) (Fig 1.1, G), externally the mouse embryo displays three branchial bars and the forelimb buds become apparent for the first time. At E10.5, the extremities and tail enlarge rapidly, most of cranial spinal ganglia are easily recognized and the first branchial bar is divided into conspicuous maxillary and mandibular processes. At E12.5 (Fig 1.1, H), the forelimb bud has visible finger rays, the somites are clearly visible from the tail to the mid-trunk region, and the eye has a four-leaf clover shape. By E14 (Fig 1.1, I), the individual fingers in the fore-foot plate are separated but not yet the toes in the hind-foot plate, also numerous hair and whisker follicles are recognizable. Just before birth, after 18.5 days of gestation, the mouse embryo has thick skin, whiskers and closed eyelids (Kaufman and Bard, 1999).

1.2 Neurogenesis and gliogenesis

As mentioned before, after the formation of the notochord and somites, the neural plate ectoderm above them begins to invaginate creating the neural tube, the embryonic precursor of the central nervous system (CNS). The differentiation of ectodermal cells to neuropithelium proceeds in a cranial-to-caudal direction. The wall of the neural tube is formed by primitive neuroepithelial cells, which proliferate enormously and develop into all types of neurons and glial cells (astrocytes and oligodendrocytes).

The lateral edge of the neural plate forms distinct cell groups, which later forms longitudinal columns on each side of the dorsal neural tube called the neural crest. The neural crest produces mesenchymal cells; neurons of the peripheral nervous system (PNS) including spinal ganglion cells and autonomic ganglion cells; Schwann cells and satellite

cells in the ganglia; connective tissue; bone; secretory cells and the outflow tract of the heart (Gammill and Bronner-Fraser, 2003).

Working out how all the neurons and glial cells are generated in the neural tube is one of the most important scientific challenges, most cells in the CNS are thought to arise from multipotential precursor cells in the neuroepithelium, the developmental potentials of which progressively restricted with time. The ventricular epithelium consists of a mosaic of precursors with diverse properties and cell fate potentials. External as well as genetic cues have a critical role in determining the fate of neuroepithelial precursor cells (Bonni and Greenberg, 1997;Bonni et al., 1997). The nature of the extracellular agents that drive precursor cells toward specific cell fates and the intracellular mechanisms by which they promote cell fate are now well understood.

During the last ten years several approaches have been taken to discover those genetic or external mechanisms that induce specific cell differentiation in the CNS. As a result, it is increasingly evident that the developing nervous system is a highly patterned tissue from a very early stage. In the anteroposterior axis, for example, there is obvious regional differentiation generating four main regions, the forebrain (prosencephalon), midbrain (mesencephalon), hindbrain (rhombencephalon), and spinal cord. In the hindbrain each of its eight enlargements, called rhombomeres, induces several tissue lineages creating several gene expression domains. One of those genes is *Krox 20*, a zinc finger protein, that is expressed in rhombomeres 3 and 5 during early development (Wilkinson et al., 1989). Later, Krox 20 can directly trans-activate the gene *Hoxb2* (Sham et al., 1993). Hox genes help to establish the anterior boundaries of odd-numbered rhombomeres, therefore

they have a very important role in specifying cell derivatives of the hindbrain (Nagy et al., 2003).

In addition, CNS dorsoventral patterning has been extensively studied in the spinal cord during development (Fig. 1.2). A model has been developed in which the forkhead domain gene, *Foxa2* (*Hnf3b*), is an important regulator of floor plate development in the mouse. It is hypothesized that Foxa2 expression in the notochord induces the expression of sonic hedgehog (Shh) (Echelard et al., 1993;Sasaki and Hogan, 1993). Shh is a secreted glycoprotein that acts in a concentration dependent manner to control the differentiation of distinct cell types in the ventral neural tube. The inductive activity of *Shh* is mediated directly even over relatively long distances (many cell diameters) (Roelink et al., 1995;Ericson et al., 1996). Combined with Shh, released by the notochord, the Bone morphogenetic proteins (BMPs) secreted from the ectoderm overlying the neural tube are thought to lead the generation of distinct classes of neurons and glial cells at specific dorso-ventral locations within the spinal cord. This pattern is also established by many transcription factor genes, such as members of the Pax, Nkx, Sox, Olig, Fox, Irx, Dbx and Lhx (Lim) families, which are differentially expressed in restricted neuroepithelial domains along the dorsal-ventral axis (Briscoe and Ericson, 1999;Rowitch et al., 2002).

The expression domains of these transcription factors delineate five progenitor cell domains in the ventricular zone (VZ) of the ventral half of the neural tube, referred to as p0, p1, p2, pMN and p3 (dorsal to ventral). Based on this, it is concluded that V0-V3 interneurons develop from the p0-p3 domains, respectively, and motor neurons (MNs) as well as oligodendrocytes (OL) arise from the pMN domain; more about this will be

discussed in the next section (Fig 1.2) (Roelink et al., 1994;Roelink, 1996;Tanabe and Jessell, 1996;Ericson et al., 1997;Richardson et al., 2000;Jessell, 2000;Briscoe et al., 2000;Lu et al., 2001).

The transcription factors can be subdivided into two functional classes based on their regulation by Shh and their interactions with each other. Class I transcription factors are repressed by Shh, whereas expression of class II factors is induced by Shh (Briscoe and Ericson, 1999;Soo-Kyung and Pfaff, 2001). Class I and class II genes have well defined boundaries of expression at the borders of the progenitor cell domains. The sharp patterns of gene expression form because the opposing class I and class II genes are incompatible in the same cell because of their cross-inhibitory activities (Soo-Kyung and Pfaff, 2001). Therefore, neurogenesis and gliogenesis depend on a hierarchy of protein-gene interactions initiated by a morphogenic gradient of Shh.

Temporal mechanisms controlling gene expression seem to operate in the progenitor domains, generating different cell types over time and increasing cell diversity within the neural tube (Marquardt and Pfaff, 2001;Soo-Kyung and Pfaff, 2001). This is particularly well characterized for the progenitors of motor neurons and oligodendrocytes. Motor neurons are among the first postmitotic cells to develop from the pMN domain, followed by a switch to the generation of OLPs (Richardson et al., 2000;Lu et al., 2000;Zhou et al., 2000;Soula et al., 2001;Zhou and Anderson, 2002). This temporal switch in specification of motor neurons to oligodendrocytes is accompanied by a change in the pattern of transcription factor expression in these cells, which may result in part from the changing

levels of Shh expressed by the notochord and floor plate during development (Sun et al., 2001;Lu et al., 2002).

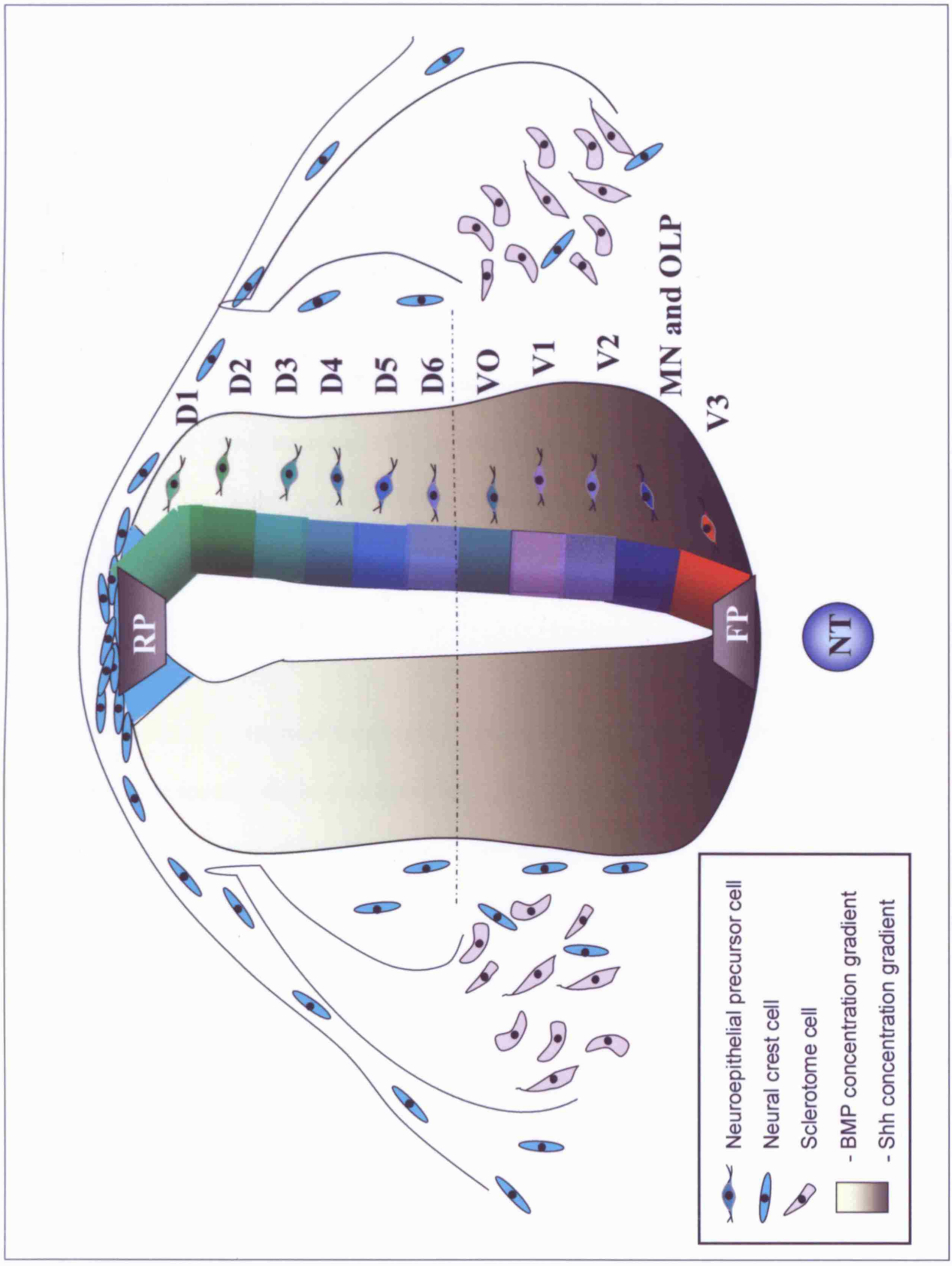


Figure 1.2. Schematic representation of the gene expression domains in the embryonic spinal cord. A transversal section of a mice embryo shows the spinal cord in the centre, the neural crest cells migrating dorso-laterally and ventrally, and the somites with already some sclerotome separating from it. The roof plate (RP) of the spinal cord releases bone morphogenetic proteins (BMP) and the floor plate (FP) and notochord (NT) release sonic hedgehog proteins (Shh) creating a concentration gradient. This gradient seems to induce the formation of eleven domains of gene expression in the spinal cord, shown as boxes of different colours. The domains in the ventral part are called, p0, p1, p2, pMN and p3 (dorsal to ventral). V0-V3 interneurons develop from the p0-p3 domains, respectively, and motor neurons (MN) as well as oligodendrocytes (OLP) arise from the pMN domain. The dorsal part is divided in six domains, D1, D2, D3, D4, D5 and D6 (dorsal to ventral). Each domain expresses particular genes throughout development.

How the combined actions of transcription factors influence their downstream target genes and ultimately specify distinct neuronal and glia cell types is not yet understood. This is one of the general aims of my Thesis work. By trying to identify regulatory elements that control *PDGFRA* expression in OLPs, I could learn something of how OLPs are specified by transcription factors and possibly by Shh.

1.3 Control of gene expression during development

Genomic DNA contains all the information required to form a living organism with all its complexity. Discovering the way gene expression is regulated in specific cell types and at

particular times during embryogenesis is still a great challenge to scientists. First of all, it is known that the regulation of gene expression in each cell can be accomplished at least at five different levels: 1) Differential gene transcription, regulating which of the nuclear genes are transcribed into RNA. 2) Selective nuclear RNA processing, regulating which of the transcribed RNAs get into the cytoplasm to become messenger RNAs. 3) Selective messenger RNA translation, regulating which of the mRNAs in the cytoplasm get translated into protein. 4) Differential mRNA stability, regulating the abundance of mRNA in the cytoplasm. 5) Differential protein modification, regulating which proteins are allowed to remain or function in the cell (Gilbert and Raunio, 1997;Gilbert, 2000).

Studies at the level of differential gene transcription have produced the first transcriptional control models for a few well studied genes, such as the human *β-globin* gene cluster, the *lysozyme* gene and some important developmental genes such as *eve-skipped* (*eve*) gene, *Myf-5*, and the *Hox* gene clusters (Dhar et al., 1989;Sham et al., 1993;Zardoya et al., 1996;Bulger et al., 1999;Santini et al., 2003;Hu et al., 2006;Ragoczy et al., 2006).

The *eve-skipped* gene (*eve*) was studied during *Drosophila* development. It is transcribed in seven bands of cells in the blastoderm stage creating a pattern of stripes. More than 800 bps upstream of the transcription start site are needed to activate the expression of the *eve* gene in all seven stripes. DNAaseI footprinting analysis showed that this region contains binding sites for activators and repressors of transcription. The activators include the transcription factors bicoid and hunchback, the repressors are giant and kruppel. Each activator binding site seems to be followed by a repressor site suggesting competitive interactions.

A different model has been discovered for the *Hox* gene clusters. *Hox* genes direct further development of tissues to form the structures and organs characteristic of each part of the body. Hox proteins are required throughout development and into adult stages. The transcription control regions of some *Hox* genes contain binding sites for their own encoded proteins creating auto-regulatory loops. In addition to these auto-regulatory loops, the chromatin structure has to be modulated by proteins encoded by polycomb-group and trithorax-group genes. Furthermore, the specificity of Hox proteins is controlled in part by the formation of heterodimers between individual Hox proteins and Exd protein in *Drosophila* or Pbx proteins in mammals. Each Hox-Exd dimer has different DNA binding specificities; in this way they control expression of different sets of genes.

Some of these studies used the strategy of creating different DNA constructs that link binding sequences surrounding the gene locus, upstream or downstream, to reporter genes like *β-galactosidase* or *luciferase*, in order to discover the location of the genetic elements regulating the expression of the gene. I also apply this strategy in this Thesis to study the transcriptional regulatory elements of the *Pdgfra* gene in transgenic mice.

1.4 Gene structure and transcriptional control elements

Every gene contains a protein coding sequence, which begins with a START codon (ATG) and concludes with a STOP codon (TAA, TAG or TGA). However, most eukaryotic genes are different from prokaryotic genes because they are not co-linear with their peptide products; that is, the gene is broken up into non-contiguous coding domains (exons)

separated by non-coding sequence (introns). The introns are included in the primary RNA transcript but are removed during subsequent RNA processing and the exons joined together (spliced) before the mRNA thus formed is transported out of the nucleus. In fact, not all of the sequence in the mRNA is protein coding, there being upstream (5') and downstream (3') non-coding (control) sequences also. In same instances these non-coding un-translated sequences are specified by separate exons.

In order to activate or inhibit a gene in specific cell types or at specific times, for example during the cell cycle or mammalian development, each gene must have regulatory elements (REs) associated with it. There are basically two main types, the *cis* and *trans*-acting regulatory elements. The *trans*-acting elements are soluble molecules that can bind the DNA of other genes, this include proteins called transcription factors, or even structural RNAs. The *cis*-acting regulatory elements are short DNA sequences where specific transcription factors or any other regulatory molecules bind. These elements act as switches to turn on or off the gene, transiently or permanently.

Two main types of *cis*-acting regulatory elements are the promoters and the enhancers (Fig. 1.3). Every gene has a promoter, which is the binding site for the basal transcriptional apparatus –RNA polymerase II and its co-factors. This provides the minimum machinery necessary to allow transcription of the gene. By convention, upstream (5') and downstream (3') directions are specified in relation to the mRNA. Thus, the promoter is upstream of the gene, near its 5'end. The term "promoter" is used in various circumstances to describe the entire upstream regulatory region of a gene or a small area around the initiation site. Basal promoters are DNA elements that interact with

the basal transcriptional machinery, usually containing a TATA box (at position -30 relative to the transcription initiation site at +1), downstream promoter element G(A/T)CG (at +30) (DPE), and/or initiator sequence (+1), at the site of transcriptional initiation. The basal promoter is not active *in vivo* unless linked to an enhancer. Not all promoters contain TATA boxes and DPEs; a recent survey indicated that approximately 14% of 205 well-characterized promoters from *Drosophila* contain both TATA and DPE elements, 55% contained only a TATA box or a DPE element, and 31% did not contain either (Kutach and Kadonaga, 2000). In vertebrate cells, house-keeping genes are often associated with basal promoter elements that are rich in CpG sequences and lack canonical TATA boxes and DPE elements.

Transcription initiates at basal promoter regions that contain conserved sequences immediately surrounding the start site. At the promoter, the RNA polymerase II interacts with a complex of over two dozen polypeptides comprising the basal transcription machinery. *In vitro*, the basal machinery, including the TATA-box binding factor TFIID, is sufficient to allow the RNA polymerase II to initiate on naked DNA templates, but *in vivo*, where the template is associated with histones (assembled into chromatin) and presumably less accessible to DNA binding factors, additional signals are required. These signals come from DNA-binding transcriptional activators bound to *cis*-elements, called enhancers, located 5' or 3' of the initiation site. Therefore, enhancer regions are found at a distance from the promoter, to either the 5' or 3'sides of the gene or even within introns. They are typically short stretches of DNA (200bp, say), each made up of a cluster of even shorter sequences (25bp, say) that are the binding sites for a variety of cell- or region-specific transcription factors. The locations of these enhancers, which could be up to 100

kb from their site of action, suggests that loops in the DNA allow these to make close contact with the basal machinery, although alternative models of action have also been suggested (Bulger and Groudine, 1999). Once bound, these transcription factor complexes interact with the basal transcriptional machinery at the promoter to enhance, or sometimes diminish, the transcription rate of the gene. Not all basal promoters have identical properties, however. Some germline-specific gene expression may involve distinct basal promoter elements and tissue-specific components of the basal transcription machinery (Hiller et al., 2001).

To summarize, enhancers are originally defined as sequences of DNA that influence transcription of genes in a distance-and orientation- independent manner. They can be regarded as platforms for the assembly of sequence specific DNA-binding transcription factors. These clusters of transcription factor binding sites can be located in both distal and proximal-promoter locations. Both activator and repressor proteins bind to enhancers, making these elements integrators of both positive and negative transcriptional information.

In conclusion, promoters and enhancers are necessary for controlling where and when a particular gene is transcribed. The communication between them can be regulated by functional properties of the basal promoter or enhancer that make certain enhancer-promoter combinations incompatible, or can be regulated by boundary elements that block enhancer-promoter interactions in a directional manner (Arnosti, 2002). These boundary elements are called insulators. DNA insulators could also prevent enhancers in one gene from inappropriately regulating a neighbouring gene.

An additional level of transcriptional control is afforded by the position of a gene relative to heterochromatic regions; most genes are located in euchromatic domains or active chromatin, and are inhibited by close proximity to heterochromatin. Constitutive heterochromatin is characterized by an abundance of repeat elements and a paucity of genes (Pirrotta, 1998). Genes in active chromatin are contained within loop structures, which are anchored at their base by the nuclear scaffold. These active chromatin domains contain three principal types of *cis*-active DNA sequences: 1) regulatory switches, like the promoters, enhancers and silencers described above. 2) regulatory "punctuation signals", like A elements located close to scaffold-matrix attachment regions (S/MARs) and 3) locus control regions (LCR) (Fig 1.3) (Sippel et al., 1992). S/MARs are double strand DNA regions that attach to scaffold or matrix material at the base of the above-mentioned loops. These attachment regions co-locate with the upstream and downstream border of the active chromatin domain. An LCR is a sequence that acts like a super-enhancer that switches the entire loop region on or off by allowing the chromatin in the loop to become in a condensed (off) or open (on) configuration (Dhar et al., 1989).

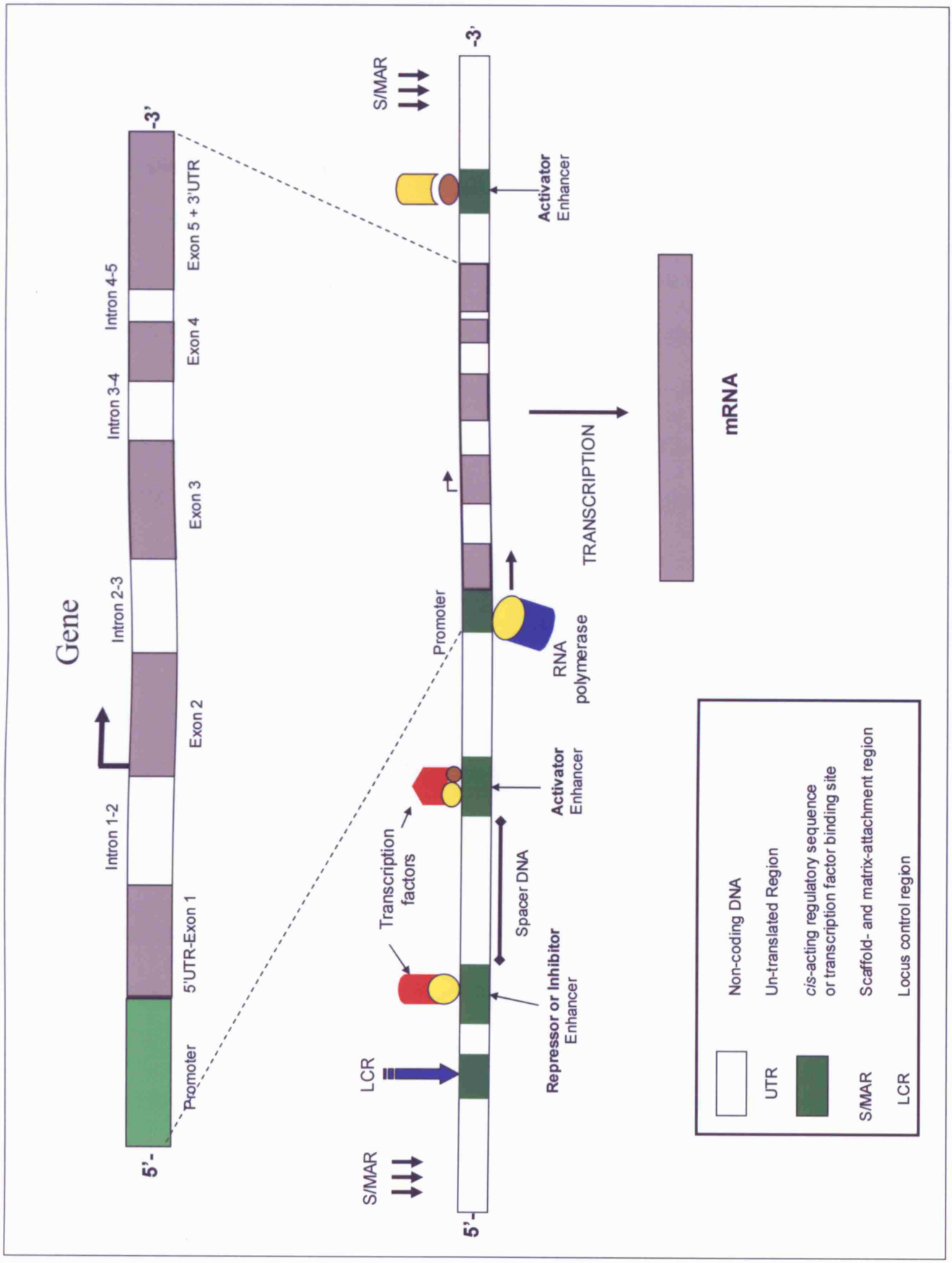

Gene
5'-
Promoter
5'UTR-Exon 1
Intron 1-2
Exon 2
Intron 2-3
Exon 3
Intron 3-4
Exon 4
Intron 4-5
Exon 5 + 3'UTR
-3'
S/MAR
LCR
Promoter
Transcription factors
Spacer DNA
Repressor or Inhibitor Enhancer
Activator Enhancer
RNA polymerase
TRANSCRIPTION
mRNA
Activator Enhancer
S/MAR
-3'
5'-
Non-coding DNA
UTR
Un-translated Region
cis-acting regulatory sequence or transcription factor binding site
S/MAR
Scaffold- and matrix-attachment region
LCR
Locus control region

Figure 1.3. Schematic representation of the basic gene structure and the location of *cis* and *trans*-regulatory elements. Exons (1 to 5) are indicated in grey boxes, including those containing 5' and 3' un-translated regions (UTR). Introns (1 to 4) are non-coding sequences shown as empty boxes in between the exons. The translation initiation site (ATG) is represented as an arrow on exon 2. *Trans*-acting regulatory elements are shown as transcription factors and RNA polymerase bound to the *cis*-acting regulatory sequences, shown as green boxes. *Cis*-acting regulatory elements are the promoter and enhancers. Enhancers could be activators or repressors of transcription. Chromatin structure elements like the scaffold and matrix-attachment regions (S/MARs), which limit the open and therefore active chromatin region, and the locus control regions (LCR), sequences that act as super-enhancers, are also shown as three black and blue arrows respectively.

A new *cis*-regulatory element has been discovered in Hox genes, it is been named global control region (GCR) because it seems to regulate the expression of clustered genes over distances of several hundred kilobases (Spitz et al., 2001;Spitz et al., 2003;Spitz and Duboule, 2005).

Eukaryotic transcriptional elements and proteins have been studied extensively in yeast, *Drosophila*, and vertebrates, and it is clear that many aspects of the transcription machinery are highly conserved, facilitating comparison among systems (Arnosti, 2002). In addition, large-scale sequencing projects have yielded comprehensive information about the protein coding capacity of entire genomes, although the transcriptional

regulatory elements are still considerably more difficult to identify from primary sequence alone. Taking all this new information in consideration, two current approaches to identify regulatory regions from genomic data involve bioinformatic methods to 1) find clusters of putative binding sites of transcription factors, and 2) phylogenetic comparisons to identify evolutionarily conserved sequences. Even though these functional genomic approaches offer us lots of clues about regulatory regions, it will still require direct experimental testing to confirm the functions of such putative regulatory regions. More about bioinformatic approaches is included in Chapter 5.

1.5 Platelet-derived growth factor receptor alpha subunit (PDGFRA)

The *PDGFRA* gene encodes a cell surface tyrosine kinase receptor for members of the platelet-derived growth factor (PDGF) family. Mammalian PDGFs have been shown to drive cellular responses including proliferation, survival, migration, and the deposition of extracellular matrix and tissue remodelling factors. Knockout studies have demonstrated that many of these cellular responses to PDGFs are essential during mouse development because the *Pdgfra* null mouse is embryonic lethal (Hoch and Soriano, 2003).

Four PDGF ligands, A, B, C, and D, have been discovered and characterized. Each ligand binds as a homodimer, but PDGFA and PDGFB can also form functional heterodimers. There are two PDGF receptors, alpha (PDGFRA) and beta (PDGFRB). They also function as homo and heterodimers. The identity of the growth factor bound to a receptor monomer

determines whether the functional receptor is a homodimer (AA, BB) or a heterodimer (AB). In vitro assays have demonstrated that ligand dimers AA, AB, CC and BB can bind to PDGFRAA (or PDGFRA); AB, CC, BB and DD bind to PDGFRAB; BB and DD bind to PDGFRBB homodimer (Betsholtz et al., 2001;Aase et al., 2002;Hoch and Soriano, 2003).

Knockout studies have demonstrated that PDGFB and PDGFRB are essential for the development of support cells in the vasculature, whereas PDGFA and PDGFRA are more broadly required during embryogenesis, with essential roles in numerous contexts, including CNS, neural crest and organ development (Yeh et al., 1993;Soriano, 1994;Leveen et al., 1994;Bostrom et al., 1996;Soriano, 1997;Fruttiger et al., 1999;Karlsson et al., 1999;Gnessi et al., 2000;Karlsson et al., 2000). In particular, PDGFRA is limiting for kidney development since mice heterozygous for the receptor gene exhibit defective kidney phenotypes. Detailed description of the PDGFRA gene expression pattern during development will be included in Chapter 3 and 4.

As a member of the tyrosine kinase receptor family, PDGFRA has two tyrosine kinase domains in the cytoplasmic region and five immunoglobulin repeats in the extracellular ligand-binding domain. When two PDGFR subunits dimerize, this activates phosphorylation of the tyrosine kinase domains which induces the binding of signalling proteins and adaptors that initiate signal transduction. PDGFRA as well as PDGFRB can activate many of the same major signal transduction pathways, including the Ras-MAPK (mitogen activated protein kinase), phosphatidylinositol 3-kinase (PI3K) and phospholipase Cγ pathways. However, there are slight differences between the proteins

that interact with the phosphorylated tyrosine kinases of each receptor, as Klinghoffer et al. (2001), showed by creating 2 complementary lines of knock-in mice in which the intracellular signalling domains of one *PDGFR* had been removed and replaced by those of the other. While both lines demonstrated substantial rescue of normal development, substitution of the PDGFRB signalling domains with those of PDGFRA resulted in varying degrees of vascular disease. Therefore, these proteins must trigger distinct signal transduction pathways that induce different functional capabilities *in vivo* (Heldin and Westermark, 1999).

1.5.1. PDGFRA discovery and gene structure description

Matsui et al. (1989) identified a genomic sequence and a cloned cDNA for a novel receptor-like gene of the PDGF receptor/CSF1 receptor family. The gene recognized a 6.4kb transcript that was co-expressed in normal human tissues with the then-known 5.3kb *PDGF* receptor mRNA. The characteristics of the new receptor were different from the previously known one (PDGFR1 or PDGFRB). This was the way PDGFRA was first described and is possible to find it named as CD140A or PDGFR2. Matsui et al. (1989) suggested that the existence of two PDGF receptors that interact in a distinct manner with 3 different PDGF isoforms, nowadays five PDGFAA, AB, BB, CC and DD, may confer regulatory flexibility in functional responses to PDGF.

Matsui et al. (1989) also localized the *PDGFRA* gene to chromosome 4q11-q12 by in situ hybridization. Disteche et al. (1989) and Gronwald et al. (1990) confirmed the assignment of the *PDGFRA* gene to 4q11-q12 by in situ hybridization and by Southern analysis of a

Chinese hamster/human cell hybrid that retained only human chromosome 4. Two years later, Hsieh et al. (1991), also assigned the *PDGFRA* gene to 4q11-q21 in the human and to chromosome 5 in the mouse by analysis of somatic cell hybrids.

The *Pdgfra* locus is closely linked to the *Kit* oncogene on mouse chromosome 5, in a region syntenic to the homologous genes in the human on chromosome 4. Therefore accumulating data through the years show that *PDGFRA*, *KIT* and *KDR* genes, all receptors of the tyrosine kinase family, are oriented head-to-tail consecutively on chromosome 4 in human and on chromosome 5 in mouse.

Kawagishi et al. (1995) isolated genomic clones encoding human *PDGFRA* and showed that the gene contains 23 exons spanning about 65 kb. The first non-coding exon is followed by a large intron of approximately 23 kb, such a large intron is also present in *CSF1R* and *KIT* genes (Roberts et al., 1988;Vandenbark et al., 1992). Exon 2 contains the translation initiator codon ATG (Fig. 1.4).

The mouse *patched* (*Pth*) mutation, which causes a white spotting pigmentation phenotype in the fur, is caused by a deletion in the *Pdgfra* gene in chromosome 5, although the extent and position of the deletion has not been precisely established (Stephenson et al., 1991;Smith et al., 1991) and the deletion might extend into a neighbouring gene(s).

In this Thesis, I report a more accurate structure and conservation of the human, mouse, chicken, xenopus and zebrafish *pdgfra* gene, using some molecular techniques like inverse

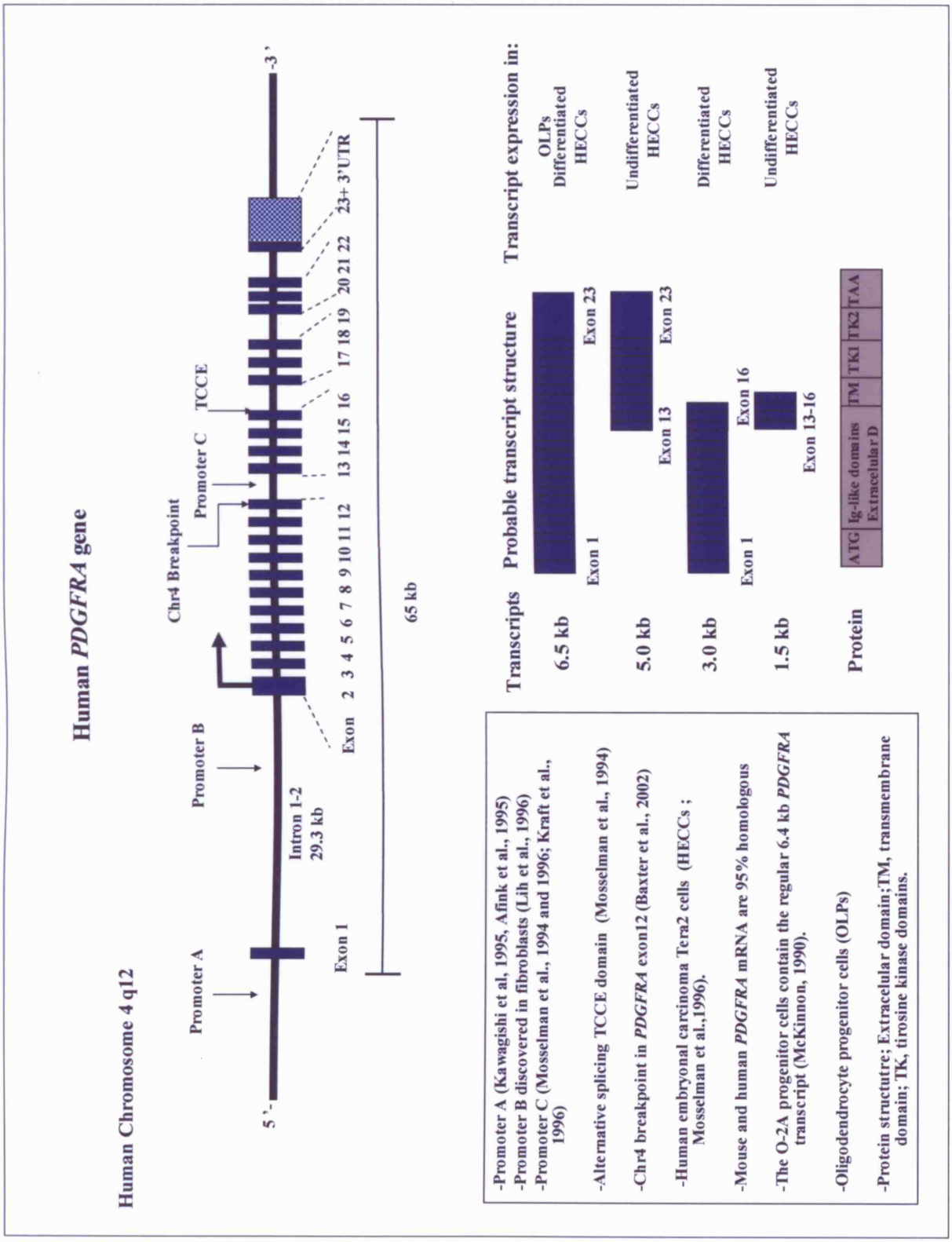


Figure 1.4. Schematic representation of the human *PDGFRA* gene, its promoters and transcripts. *PDGFRA* gene has 23 exons. Exons are represented as blue boxes. Intron

1 is ~29.3 kb long. The ATG site is found in exon 2. Three promoters are described, one before exon 1 (A), the second before exon 2 (B) and the third before exon 13 (C) (Mosselman et al., 1994;Afink et al., 1995;Lih et al., 1996;Mosselman et al., 1996;Kraft et al., 1996). Chromosome 4 breakpoint was found in exon 12 (Baxter et al., 2002). Four transcripts are described: 6.4kb, 5 kb, 3 kb and 1.5 kb long. Transcripts 6.4 and 3 kb were found in differentiated human embryonal carcinoma Tera2 cells (HECCs), and transcripts 5 kb and 1.5 kb in undifferentiated HECCs (Mosselman et al., 1996). In addition, oligodendrocyte O-2A progenitors express the 6.4 kb transcript (McKinnon et al., 1990). An alternative splicing TCCE domain was found in exon 15 and is thought to help for the formation of 3kb and 1.5kb transcripts (Mosselman et al., 1996). PDGFRA protein structure consists of extracelular domain with five Ig-like loops, transmembrane domain and two tyrosine kinase domains.

PCR in case of the human gene (Chapter 3), but mainly analysing the new releases in genome browser websites of each vertebrate genomic sequence (Chapter 5).

1.5.2. PDGFRA in spinal cord and brain development

As previously mentioned (section 1.2.1), it was discovered in the early 1990's that *Pdgfra* is expressed by oligodendrocyte progenitors (OLPs) in the CNS of the rat (Pringle et al., 1992;Yeh et al., 1993;Pringle and Richardson, 1993;Hall et al., 1996). Nowadays, it is known that at E12.5 in mice, E14 in rat, E6 in chick and stage 40 in xenopus a few cells in the ventral neuroepithelium start expressing *pdgfra* (Pringle et al., 1992;Mora, 2005). These cells are located in the pMN domain of the ventral spinal cord, where motor

neurons are generated earlier (Yu et al., 1994). Later during development, *Pdgfra* positive progenitor cells seem to migrate widely throughout the spinal cord until both white and grey matter contains many distributed cells (Pringle and Richardson, 1993;Pringle et al., 1998).

Pdgfra is detected in OLPs but not in mature oligodendrocytes (OLs) (Hart et al., 1989;Pringle and Richardson, 1993). Furthermore, OLPs and newly differentiated oligodendrocytes are the only cells that express *Pdgfra* in the rat optic nerve (Hart et al., 1989). In addition, analysis of PdgfA knockout mice show a severe reduction in the number of OLPs and OLs in the spinal cord, which implies that PdgfA, acting through Pdgfra, is crucial for OLP proliferation *in vivo* (Fruttiger et al., 1999).

Pdgfra positive cells in the late embryonic rat spinal cord are believed to correspond closely to so-called 0-2A progenitors, first characterized in cultures of rat optic nerve. If Pdgfra-positive cells were immunoselected with an anti-Pdgf-A antibody and cultured *in vitro*, they behaved in a similar way as 0-2A progenitors. For example, when cultured under appropriate conditions they would differentiate into either OLs or type-2 astrocytes. Conversely, when Pdgfra positive cells were selectively removed from cultures of rat spinal cord cells by complement-mediated cell lysis, then oligodendrocyte production was strongly inhibited (Hall et al., 1996).

The ventral origin of OLPs in the spinal cord and telencephalon seems to be due to the signals provided by Sonic hedgehog (Shh). As mentioned before, Shh is expressed first in the notochord, which induces the formation of the floor plate in the early developing

neural tube. Later the floor plate becomes the major source of Shh, creating a high-low concentration gradient of Shh from ventral to dorsal in the spinal cord. Part of the function of Shh to promote oligodendrocyte generation seems to be through the induction of *Olig* genes which encode basic helix-loop-helix (bHLH) transcription factors (Lu et al., 2000;Nery et al., 2001). *Olig* genes on their own can not substitute Shh, so Shh must also activate additional genes that cooperate with Olig to promote oligodendroglial fate (Park et al., 2002).

In addition, the generation of Pdgfra positive-OLPs from the ventral VZ requires a defined concentration of Shh (Pringle et al., 1996;Orentas et al., 1999;Soula et al., 2001). However, it now seems likely that Pdgfra positive-OLP generation is not uniquely dependent on Shh, because some OLPs are now known to arise in the dorsal spinal cord, where they are likely to be specified by signals other than Shh (Fogarty et al., 2005;Vallstedt et al., 2005;Cai et al., 2005).

No direct or indirect interaction between Shh signalling and the activation of *PDGFRA* gene expression in OLPs has been discovered yet. One of the main links could be through the intracellular transducers of Shh, which are the Gli1, Gli2 and Gli3 proteins (Altaba, 1999a;Altaba, 1999b). They mediate different aspects of Shh signalling and have both complementary and antagonistic activities (Park et al., 2000;Dahmane et al., 2001;Qi et al., 2003;Sanchez et al., 2004;Altaba et al., 2004;Stecca and Altaba, 2005;Stecca et al., 2005;Clement et al., 2007). All three are expressed in the ventral neural tube. The most recent studies concerning the relationship between Gli transcription factors and *PDGFRA*

gene activation was by Xie et al.(2001) using cancer cells; this research will be described in the next section.

Transcription factors other than the Glis could activate or repress the expression of several genes like *PDGFRA* within pMN precursors. An example of these interactions is that involving Olig 2 and Nkx 2.2 (Zhou et al., 2001;Sun et al., 2003). Olig 2 interacts with Nkx 2.2 at two distinct stages. During neurogenesis, a cross-repressive interaction appears to establish the precise boundary between p3 and pMN domains. At later times, a cooperative interaction is likely, because Nkx 2.2 promotes maturation of OLPs specified by the expression of Olig 2. Apparently, Olig 2 can form a specific physical complex with Nkx2.2 protein but not with Nkx6.1. Therefore, although the Olig2-Nkx2.2 interaction might be important later, to establish the pMN-p3 boundary and the maturation of OLPs, it is still insufficient for the initial specification of OLPs in developing spinal cord (Sun et al., 2003). This interpretation is consistent with the finding that $Pdgfra^{+}$ OLPs form at the right time and place in the ventral VZ of *Nkx 2.2* null mice, but do not mature into differentiated OLs in the spinal cord white matter, according to MBP and PLP expression pattern analysis (Qi et al., 2001).

1.5.3. PDGFRA in cancer and its possible activation by the sonic hedgehog signalling pathway

Considerable insight into the role of the sonic hedgehog pathway in human cancers and vertebrate development came from the discovery that mutations in *patched* (*Pth*) mice are associated with basal cell nevus syndrome (BCNS), an autosomal dominant disorder

combining developmental anomalies and tumors, particularly basal cell carcinomas (BCCs). Sporadic BCCs, the most common human cancer, consistently have abnormalities in the hedgehog pathway, and often mutations in *Pth*. In addition, somatic mutations in smoothened (smo), another protein in the hedgehog pathway, occur in sporadic BCCs. The downstream molecule Gli1 is known to mediate the biological effect of the hedgehog pathway and is itself upregulated in all BCCs (Altaba, 1999a;Dahmane et al., 2001;Altaba et al., 2002;Sanchez et al., 2004;Altaba et al., 2004;Stecca et al., 2005;Sanchez and Altaba, 2005;Clement et al., 2007). Gli1 can drive the production of BCCs in the mouse when overexpressed in the epidermis. Xie et al. (2001) showed that Gli1 can activate *PDGFRA* and that functional upregulation of *PDGFRA* by GLI1 is accompanied by activation of the Ras-ERK pathway, which is associated with cell proliferation. The relevance of this mechanism in vivo is supported by a high level of expression of PDGFRA in BCCs in mice and humans (Fig.1.4). From these and other observations, Xie et al. concluded that increased expression of the *PDGFRA* gene may be an important mechanism by which mutations in the hedgehog pathway cause BCCs.

PDGFRA gene amplification occurs in malignant gliomas, and a 2.1 kb deletion of the human *PDGFRA* gene in 1 of 9 gliomas was described, deleting residues 375-455 from the cDNA, which correspond to 81 aminoacids and removes a part of the fourth and fifth immunoglobulin-like domain in the extracellular region of the receptor (Fleming et al., 1992;Kumabe et al., 1992;Hermanson et al., 1996). Many human gliomas express high levels of *PDGFRA* together with its ligands, thus creating putative autocrine growth factor loops within these tumors (Nister et al., 1991;Hermanson et al., 1992). A large subgroup of gliomas over-expresses *PDGFRA* at the mRNA level (Hermanson et al., 1996).

PDGFRA expression as well as *NG2* expression has been described in 100 % of the tested oligodendrogliomas, three out of three pilocytic astrocytomas, and one of five glioblastoma multiform (Di Rocco et al., 1998;Shoshan et al., 1999). These findings, together with the observation that a truncated PDGF receptor inhibits glioma cell proliferation (Strawn et al., 1994), strongly suggest a prominent role for the PDGFRA in glioma development. Therefore, PDGFRA has been considered a promising drug target for glioma.

1.5.4. *PDGFRA* fusion genes in leukaemia and hypereosinophilic syndrome

Chronic myeloid leukemia (CML) is characterized by the presence of the BCR/ABL fusion gene, usually in association with a t(9;22) (q34;q11) chromosomal translocation. Baxter et al. (2002) reported the identification and cloning of a rare variant translocation, t(4;22) (q12;q11), in 2 patients with a CML-like myeloproliferative disease. An unusual in-frame BCR/PDGFRA fusion mRNA was identified in both patients, with either BCR exon 7 or exon 12 fused to short BCR intron-derived sequences, which were in turn fused to part of *PDGFRA* exon 12. Sequencing of the genomic breakpoint junctions showed that the chromosome 22 breakpoints fell in BCR introns, whereas the chromosome 4 breakpoints were within *PDGFRA* exon 12 (Fig. 1.4).

To emphasize the importance of any alteration of the *PDGFRA* gene in human diseases, Cools et al. (2003) demonstrated that idiopathic hypereosinophilic syndrome is often caused by an interstitial deletion on chromosome 4q12 resulting in fusion of *PDGFRA* and

FIP1L1, a neighbouring gene. The *PDGFRA-FIP1L1* gene is a constitutively activated tyrosine kinase that transforms hematopoietic cells. This tyrosine kinase can be inactivated by the drug imatinib, causing clinical remission.

1.5.5. *PDGFRA* in spina bifida: animal models and genetic variability.

Helwig et al. (1995), reported that mice that are doubly heterozygous for the mutants *undulated* and *patched*, mice with mutations in the *Pax1* and *Pdgfra* gene respectively, have a phenotype reminiscent of an extreme form of spina bifida occulta in humans. The unexpected phenotype in double –mutant and not in single –mutant mice showed that novel congenital anomalies such as spina bifida can result from interaction between products of independently segregating loci. This is an example of digenic inheritance.

Mouse models indicated that deregulated expression of the *Pdgfra* gene causes congenital neural tube defects (NTDs), and mutant forms of PAX1 that have been associated with NTDs caused deregulated activation of the human *PDGFRA* promoter. Joosten et al. (2001), identified 5 different haplotypes in the human *PDGFRA* promoter, of which the 2 most abundant, designated H1 and H2-alpha, differ in at least 6 polymorphic sites. In a transient transfection assay in cultured human osteosarcoma cell lines, the 5 haplotypes differed strongly in their ability to enhance reporter gene activity. In a group of patients with sporadic spina bifida, haplotypes with low transcriptional activity, including H1, were underrepresented, whereas those with high transcriptional activity, including H2-alpha, were overrepresented. When testing for haplotype combinations, H1 homozygotes

were fully absent from the group of both sporadic and familial spina bifida patients, but strongly underrepresented in unrelated controls. The data indicated that specific combinations of naturally occurring *PDGFRA* promoter haplotypes strongly affect NTD genesis.

1.5.6. Previous research into *PDGFRA* transcriptional gene regulation

The cloning and characterization of the 5' flanking regions of the human, mouse and rat *PDGFRA* gene by *in vitro* experiments have helped to find *cis*-regulatory elements and transcription factors involved in *PDGFRA* gene regulation (Wang and Stiles, 1994;Afink et al., 1995;Kitami et al., 1995;Joosten et al., 2002).

Initially, Wang discovered that retinoic acid (RA) and dibutyryl cyclic AMP (cAMP) stimulated transcription of the *PDGFRA* gene in embryonal carcinoma cells (line F9). Processed mRNA transcripts appeared within 4 hr after exposure to these agents, and functional alpha: alpha homodimers appeared within 24 h (Wang et al. 1990). Subsequently, a partial genomic map that locates the promoter and transcription start sites of the mouse *Pdgfra* gene was constructed by Wang and Stiles (1994). They isolated genomic clones that encompass the 5' end of the *Pdgfra* mRNA transcript and extend 10 kb into the upstream flanking region of the gene. After transfecting their DNA constructs into bladder epithelial cell lines, one of them contained a RA and cAMP response element between -912 and -447 upstream of the transcription start site, which drives expression of the reporter gene selectively in cells that express *Pdgfra*.

Subsequently, Afink et al. (1995) showed that *PDGFRA* promoter activity in human teratocarcinoma cells is strongly up-regulated following RA and cAMP treatment. They tested a panel of 5'-deletion mutants of the *PDGFRA* promoter cloned into luciferase reporter gene vectors. Their analysis revealed that *in vitro* the promoter fragments can drive the reporter gene expression in carcinoma cells, that the region from -441 to +118 is sufficient to establish high level *PDGFRA* promoter activity, and specifically, that this increase is controlled by a 23 bp ATTA-sequence-containing element (named parATTA) located in the -52/+118 region of the promoter. Later, Joosten et al. (2002) using electrophoretic mobility shift assays (EMSAs) showed that multiple transcription factor complexes can compete for binding to the parATTA sequence. One of those complexes is the one form by the transcription factors PRX and PBX. Since this PRX-PBX complex is present in both, undifferentiated teratocarcinoma cells (Tera-EC) and differentiated teratocarcinoma cells with neuronal and endodermal characteristics (Tera-RA), RA/cAMP-induction of *PDGFRA* promoter activity seemed to require an additional activation step. To test this, they did point mutations in the parATTA element, which reduced binding affinity for PBX-PRX proteins in differentiated Tera-RA cells, and as a consequence the activation of the *PDGFRA* promoter by RA/cAMP induction. Therefore, they concluded that the parATTA element is essential for normal *PDGFRA* transcriptional regulation and for its activation upon RA/cAMP treatment (Joosten et al., 2002).

Another example of these *in vitro* experiments analysing the *PDGFRA* regulation is the study of Helwig et al. (1995) who suggested a functional relation between *PDGFRA* and the homeodomain transcription factor PAX1. Three years later, Joosten et al. (1998)

showed that PAX1 acts as a transcriptional activator of the *PDGFRA* gene in differentiated human embryonal carcinoma cells, by using the human *PDGFRA* promoter linked to a luciferase reporter.

Several other extracellular factors have been described that affect *PDGFRA* expression at the mRNA and protein level, including transforming growth factor-β (TGF- β) (Yamakage et al., 1992), interleukin-1 β (IL-1 β) (Lindroos et al., 1997;Bonner et al., 1998), tumor necrosis factor α (TNF-α) (Battegay et al., 1995), and basic fibroblast growth factor (bFGF)(McKinnon et al., 1990;Bonner et al., 1996).

To identify *cis*-acting regulatory elements that dictate expression of the *PDGFRA in vivo*, Reinertsen et al. (1997), generated transgenic mice bearing the reporter gene *β-galactosidase* (*lacZ*) under the control of a 6 kb promoter sequence. Expression of *lacZ* was monitored throughout embryonic development, with special focus on nervous tissue, skeleton, and several organ systems wherein *PDGFRA* expression is thought to play a pivotal role. In several independent transgenic mouse strains, *lacZ* expression recapitulated predominant features of *PDGFRA* gene expression during mouse development. However, it did not express in OLPs at any developmental stage. These results demonstrate that certain critical tissue specific regulatory elements for *PDGFRA* expression are located within a 6 kb upstream region of the *PDGFRA* gene.

Another analysis of the *PDGFRA* upstream sequence was done by (Zhang et al., 1998). The purpose of their study was to demonstrate the *in vivo* promoter function of genomic DNA fragments representing the 5'-flanking part of the human *PDGFRA* gene. *PDGFRA*

promoter fragments of 2.2, 0.9 and 0.4 kb, ligated to a *LacZ* reporter gene, were microinjected into fertilized mouse eggs and transgenic mouse lines were established. The expression patterns were basically similar in the 2.2 and 0.9 kb lines and overlapped grossly the endogenous *PDGFRA* gene expression pattern. The transgenic line with the highest expression level was chosen for detailed analysis. Expression was mainly confined to tissues of mesodermal and neural crest origin. No expression was found in epithelial tissues of endo- or ectodermal origin. These promoter fragments were also active in neuroepithelium and in certain CNS neurons, which did not normally express *PDGFRA* mRNA, but all failed to specify reporter expression in OLPs. Thus, the isolated human *PDGFRA* 2.2 kb promoter contains some but not all of the regulatory elements that are necessary to establish tissue specific gene expression during development.

Subsequently, Sun et al. (2000) generated transgenic mice by pronuclear injection of a 380 kb yeast artificial chromosome (YAC) containing the human *PDGFRA* gene. In contrast with previous plasmid-based transgenes containing only upstream *PDGFRA* control sequence (Reinertsen et al., 1997;Zhang et al., 1998) this YAC transgene was reported to have expressed faithfully in OLPs and therefore it seems to contain the specific cis-regulatory elements to express the *PDGFRA* gene in OLPs. In addition, it was also expressed in neural crest cells and sclerotome. Furthermore, this YAC was shown to rescue the profound craniofacial abnormalities and spina bifida observed in the *PDGFRA* knockout mice. However, these *PDGFRA* knockout/YAC mice die after birth due to respiratory failure caused by a defect in prenatal lung growth. Analysing sections of lung tissue, Sun et al. (2000) found that lung smooth muscle cells progenitors did not express

the *PDGFRA* YAC transgene and possibly this was the root cause of failure of lung growth and ultimately perinatal death.

1.6 Aims of this thesis

In this PhD thesis, I present *in vivo* and *in silico* evidence concerning the organization of transcriptional control elements in and surrounding the *PDGFRA* gene. I focus on the activation of *PDGFRA* gene transcription in specific cell types during development.

My experimental approach was yeast and bacterial artificial chromosome (YAC and BAC) transgenesis and bioinformatical analysis.

My particular goal was to narrow down the region where the specific regulatory element for OLPs could be located. Analysis of this element could eventually lead to the identification of transcription factors that activate *PDGFRA* expression specifically in OLPs and therefore provide insights into the downstream readouts of Shh, RA and/or FGF signalling. Furthermore, the discovery of this specific OLP regulatory element could provide a powerful tool to drive any transgene expression only in OLPs. The long-term goal will be to use these SREs as tools for genetic manipulation of these cells in the embryonic and adult CNS.

Using an *in vivo* YAC/BAC transgenic analytical approach, and an *in silico* bioinformatical comparative analysis of non-coding DNA sequences of several vertebrate species, I narrow down the search for SREs of the *pdgfra* gene.

Chapter 2

Materials and Methods

2.1 Bacteriology

2.1.1. Bacterial strains, growth and storage.

The electro-competent Escherichia coli strain XL1-Blue (*supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 lac⁻ F'* [proAB$^+$ lacI^q LacZΔM15 Tn10 (tet^l)]) was used to carry out most of the plasmid cloning. These bacteria were grown at 37°C on LB-agar plates. If selecting with ampicillin, this was added to the LB media or molten LB-agar (after cooling to 55°C) at a final concentration of 50-100 μg/ml. Ampicillin was diluted in ultra-pure water at a concentration of 100 mg/ml stock, filtered with a 0.22 μm filter and stored in aliquots at -20°C. Liquid cultures were agitated at 250 rpm and incubated overnight at 37°C.

Long term storage of bacterial strains was done by adding glycerol to overnight cultures at a final concentration of 15% (v/v). These stocks were stored in 1 ml aliquots at -20°C.

2.1.2. Transformation of competent bacteria.

Plasmid DNA was dried on a circle of filter paper, stored at room temperature and kept away from dust. In order to resuspend the plasmid DNA again, the circle of filter paper

was cut out and put into a 1.5 ml eppendorf tube with 20µl of destillated water. After 5 min, the tube was spun for 1 min at 13 000 rpm.

Once the plasmid DNA was diluted into water, 3 µl of this solution was added into an ice cold 2 mm electroporation cuvette. The electro-competent bacteria XL1 Blue was thawed slowly on ice (2.1.1). 20µl of these bacteria was added into the electroporation cuvette and mixed carefully with the DNA solution. Then the cuvette was dried and set into the Micropulser BIO-RAD. The program Ec-2 for bacteria was set and one pulse was given to the bacteria. In consequence, the bacteria have now pores in their membrane where the foreign plasmid DNA can penetrate and get in the bacteria creating a new strain. Immediately after the pulse, the sample was put on ice and after waiting 30 seconds, 100 µl of LB media were added slowly. The bacteria were then transferred to a 1.5 ml eppendorf tube and incubated at 37°C for 15 min. Only one drop of this solution was needed to streak on a 50 µg/ml ampicillin-LB plate, in case the cloned plasmid contained the ampicillin gene otherwise another selection antibiotics could be added. The plate was then incubated at 37°C overnight. The rest of the bacterial solution was stored in the -20°C freezer.

12 colonies from the ampicillin-LB plate were picked and incubated in 5 ml of 50 µg/ml ampicillin-LB media at 37 °C, agitating at 250 rpm, overnight. Once the colonies had grown, the isolation of the plasmid DNA was done by alkaline lysis (miniprep, 2.2.1.). In order to detect the colony that contains the experimental plasmid, a restriction enzyme assay was done or alternatively a PCR of a known plasmid DNA sequence (2.2.8 and 2.2.14).

2.1.3. β-galactosidase selection system of bacterial strains.

When cloning into bacterial vectors containing the *LacZ* reporter gene, a blue/white color screen or β-galactosidase selection could be used to distinguish between colonies that contain vectors with or without inserts. Depending of the design of each vector is the color of insert-containing colonies. For example, for some common vectors colonies containing re-ligated vectors are blue after the β-galactosidase reaction, whereas colonies containing vectors with inserts are white because of the disruption of the *LacZ* gene. In this case each agar plate was spread with 20µl of 50mg/ml X-gal (5-bromo-4 chloro-3 indolyl-β-D-galactopyranosidase) and 100µl of 100mM IPTG (isopropyl B-D-thiogalactopyranosidase), and then incubate it at 37°C for 30 min. After incubation, a small drop of transformed bacteria were spread on the XGal/IPTG plates and incubated at 37°C overnight.

2.2 Molecular Biology

2.2.1. Preparation of plasmid DNA by alkaline lysis (miniprep)

The electroporated bacteria were incubated in 5 ml of 50 µg/ml ampicillin- LB media to select the ampicillin resistant colonies. 1 ml of culture was taken in an eppendorf tube, spun 6500 rpm for 1 min. Later, the supernatant was thrown away and the pellet re-dissolved in 100 µl of ice cold Solution I (25mM Tris-HCl pH8, 50mM glucose, 10mM EDTA), this solution was vortex only 10 seconds. Fresh Solution II (0.2 M NaOH, 1%

SDS) was prepared. 200 µl of solution II were added and mixed by inverting six times gently, and then the samples were incubated on ice for 5 min. Similarly, 150 µl of Solution III (3M potassium acetate, 2M acetic acid) were added and then mix well again but this time the samples were kept at room temperature during 5 min. Later, the samples were spun at 10 000 rpm by 5 min and the supernatant was taken carefully without contaminating it with any proteins or fat floating in the sample. Next, 1ml of isopropanol was added, and again the samples were vortex and spun at 6500 rpm by 5 min at room temperature. This time the supernatant was removed carefully in order to keep the pellet of DNA at the bottom of the eppendorf tube. To wash the pellet, 500 µl of 70% ethanol were added, vortex and spun 1 min at 6500 rpm twice, then the pellet was let to dry. Lastly, the DNA pellet was dissolved in 100 µl of destilated water by first waiting 2 min at room temperature and then mixing with vortex.

In order to analyse this DNA, 7µl of DNA solution extracted through this miniprep procedure were used for restriction enzyme assay (2.2.8) and only 0.5 µl or less for PCR (2.2.14).

2.2.2. Isolation of genomic DNA for mice genotyping

3-5 mm of the tail tips of two-week old mice were cut and put in an eppendorf tube. 500 µl of Genomic Prep Buffer (50mM Tris-HCl pH 8, 100mM EDTA, 100mM NaCl, 1% (w/v) SDS) and 12 µl of 10µg/µl proteinase K (Sigma) were added per tail. The samples were mixed and incubated at 55°C, overnight.

Next day, the degraded tissue was vortex vigorously, then 200 µl of 6M ammonium acetate was added, vortex again and chilled on ice for 10 to 15 minutes. This solution was centrifuged at 13000 rpm by 10 minutes at 4 °C. The supernatant was kept and 500 µl of ice cold isopropanol was added to it, mixed again and incubated 10 minutes on ice. Later, the solution was centrifuged again at 13000 rpm for 10 minutes at 4°C. The supernatant was discarded, but the DNA pellet was washed twice with 200 µl of 70% ethanol, vortex and spun 1 min at 6500 rpm, and then let to dry. Finally, the DNA was re-suspended in 100 µl of ultra-pure water. This DNA was later used for genotyping each mice by PCR (2.2.14) and/or Southern blot (2.2.12).

2.2.3. Extraction of DNA with phenol/Chloroform/iso-amyl alcohol

DNA samples to be extracted or purified from protein contamination were added 0.5% SDS and 20 µg of Proteinase K (Sigma), mixed and incubated at 55°C for 15 min. After the protein digestion, an equal volume of 25:24:1 (v/v) phenol/chloroform/iso-amyl alcohol was added to the solution. The sample was vortex and centrifuged at 10 000g for 5 minutes. The upper aqueous phase was collected and re-extracted twice and then with an equal volume of 24:1 (v/v) chloroform/iso-amyl alcohol; then vortex, centrifuged and collected again as before. This DNA solution can then be precipitated to finally get the pure DNA. The phenol stock used was buffered in Tris-HCl pH 7.5 (Sigma) and stored in aliquots at -20°C.

2.2.4. Precipitation of DNA

There were two methods to precipitate DNA. In the first one, DNA samples to be precipitated were made 0.3 M NaOAc and 2.5 volumes of 100% ethanol added, then the mixture was vortexed, incubated at -20°C for 30 minutes or left overnight, and centrifuged at 6500 rpm for 10 min at 4°C. If less than 1-2μg of DNA or >200 bp DNA fragments were to be precipitated, then 10 μg of glycogen was added to the DNA solution as a carrier during precipitation. The DNA pellet was washed with 100 μl of 70% (v/v) ethanol, centrifuged at 6500 rpm for 10 min at 4°C, air dried and resuspended in 50-100 μl of TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA) or ultrapure water. The second method to precipitate the DNA is to add isopropanol to the DNA solution, mix and incubate at -20°C for 30 min at least, then centrifuge and wash the pellet with ethanol as previously described.

The DNA concentration was calculated by UV spectrophometry quantification.

2.2.5. Agarose gel electrophoresis of DNA

The agarose (Multi-Purpose, molecular grade from BIOLINE) was used for all gel electrophoresis from 0.8% to 2.5% (w/v). The percentage of the agarose gel depends on the size of the DNA that wants to be analysed.

1X TAE (0.04M Tris-acetate pH7.5, 1mM EDTA) was always the buffer used to prepare gels and run most of the electrophoresis. The concentration was changed to 0.25X TAE to run long sequence DNA (between 40-500 kb) on a minigel or PFGE (2.4.5.1).

The agarose was dissolve in the appropriate volume of TAE buffer by boiling in a microwave oven and let to cool to <60°C approximately. Ethidium Bromide (10 mg/ml stock) was added into the agarose to a final concentration of 0.2 μg/ml. The agarose was poured into a suitable size electrophoresis gel container (Horizon electrophoresis system, Life Technologies, Ltd) and let to set. The DNA samples were then mixed with 1X DNA loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF and 15% (w/v) Ficoll (Type 400, Pharmacia)) and loaded in the wells of the agarose gel. The gel electrophoresis was run at 5 V/cm (Horizon electrophoresis system, Life Technologies, Ltd).

The DNA was then visualised on an ultraviolet (UV) transiluminator within the MultiImage Light Cabinet (Alpha Innotech Corporation). A picture was taken, controlling the exposure and contrast with the Alpha Imager 1220 5.5v software program, and printed in a Digital Graphic printer.

2.2.6. Gel purification of DNA using the Geneclean II kit (Bio 101).

DNA bands were purified from a gel by two different techniques depending on the amount of DNA available and the purification demands. The first technique uses the Geneclean II

kit (Bio 101). After electrophoresis, DNA bands were cut out of the gel (approximately 0.5mm X 0.8mm agarose blocks) and collected in an Eppendorf tube. Three volumes of 6M Sodium iodide was added, assuming a gel density of 1g/ml, and then incubated at 55°C for 5 minutes, tap a bit the solution to check if the gel melted completely, alternatively incubate another 2 min, take out the eppendorf and chilled on ice. Frequently, adding only 5 µl of supplied silica matrix will be enough to bind most of the DNA in a band. The DNA/silica matrix solution was vortex thoroughly, incubated on ice for 5 minutes, spun a few seconds, washed 3 times in the supplied wash buffer and re-suspended in TE buffer. In order to separate the DNA from the silica matrix, the solution was incubated at 55°C for 5 minutes, spun again and the supernatant was recovered containing the pure DNA. This method was recommended for bands containing lots of DNA, because the concentration of DNA recovered were normally below the expected. This normally due to avoid any silica matrix contamination in the last DNA sample, because it is known to affect during other cloning procedures.

DNA purification kits using microcolumns (Qiagen) were more efficient to purify bands with low concentrations of DNA. The resulting pure DNA was then used for further cloning straight away.

2.2.7. Gel purification of DNA by electroelution into dialysis bags.

The DNA gel band, containing the selected DNA to be purified from the agarose, was cut out and placed in a Petri dish. A piece of dialysis tubing (selected according to the size of

the DNA to be purified, stored in EDTA buffer at 4°C and prepared as described by (Sambrook et al., 1989) was cut to size and rinsed with distilled water to remove completely the EDTA buffer. Over a tray, one end of the tubing was clipped and the dialysis tube was first rinsed and later filled with 1X TAE buffer, in order to put the gel band inside. All the bubbles and some of the 1X TAE were removed, and then the clip at the other end was closed. Later, the dialysis tube with the DNA band inside was placed floating in a gel tank full of 1X TAE buffer, the electrophoresis chamber was closed and run at 5 V/cm for a few minutes. The time of electrophoresis depended on the size of the agarose band. It was continued until the DNA was out of the gel; this was checked by looking at the already EtBr-stained DNA under a UV transiluminator. Once the DNA was completely out of the gel, the dialysis tubing was placed inverted in the electrophoresis tank and the power turned on only 30 secs. This step separated the DNA from the edge of the dyalisis tubing. Finally, one clip of the dyalisis tubing was opened and the 1X TAE buffer, now containing the DNA, was carefully recovered.

2.2.8. Enzymatic DNA digestion and DNA ligation reaction.

Restriction enzyme digestions were carried out considering the enzyme manufacturers protocols (New England BioLabs Ltd). Approximately 10 μg of DNA was digested in 1X restriction buffer and approximately 20 units of restriction enzyme, depending on its efficiency. The solution was then mixed, centrifuged a few seconds and incubated at 37°C overnight or after at least two hours in a water bath.

Ligation reactions were normally done using the enzyme T4 ligase and T4 ligase buffer (New England Biolabs Ltd.). The DNA concentration of the fragments to be ligated varied according to the experiment, normally insert: vector molar ratio of 3:1. This could be estimated with the equation X = ((50/V) x I) x3, where X is the amount of insert in ng, V is the length of the vector in base pairs and I is the length of the insert in base pairs. The concentration of the enzyme, buffers and ATP was normally 1µl T4 ligase, 1 µl 10X T4 buffer and 0.5 µl 1mM ATP in a 10 µl total volume reaction, incubated at 16°C, overnight in a MWG thermocycler for best results.

2.2.9. Random-primer labeling of DNA probes with [α-^{32}P] dCTP.

50 ng of linear DNA template was resuspended in ultrapure water in order to get 34 µl of total volume. This DNA was then denatured at 95°C for 5 min, and chilled on ice for two minutes. After that, 10µl of 5X Labelling buffer, 2 µl of 25X dNTPs, 4 µl of [^{32}-P] dCTP (10µCi/µl, Amersham), and 0.5 µl of large fragment DNA polymerase I (Klenow), were added and mixed thoroughly by pipeting up and down. All the precautions to work with radioactive materials were followed. The radioactive reaction was then incubated at room temperature by 2 hrs. After that, the DNA polymerase reaction produced enough [^{32}-P] dCTP-labelled probe. The radiolabelled probe was filtered over a Micro Bio-spin 30 column (BioRad) to remove unincorporated radionucleotides (2.2.11). For a Southern blot, this probe needs to be denatured at 95°C for 5 min before adding it to the hybridisation buffer (2.2.12).

2.2.10. 5'termini labeling of oligonucleotides probes with [γ-^{32}P] ATP

On some occasions, small specific oligonucleotides were radiolabelled to use as probes for a Southern blot. 50 to 100 ng of oligonucleotides were mixed with the following reagents. 1µl of 10X T4 polynucleotide kinase buffer (400mM Tris-HCl pH7.5, 100mM $MgCl_2$ and 50 mM DTT), 4 µl of [γ–^{32}P] ATP (10 µCi/µl), 3 µl dH_20, 1µl T4 polynucleotide kinase (10 units, Promega). This reaction was then incubated at 37°C for 45 minutes, inactivated at 65°C for 10 minutes, filtered over a MicroBio-Spin (BioRad) (2.2.11), and finally denatured by heating to 95°C for 5 minutes before adding it to the hybridization buffer for the Southern blot (2.2.13).

2.2.11. Purification of radioactive probes.

The radioactive probe was filtered in a Micro Bio-Spin P-30 Tris Chromatography column (BIO-RAD) in order to eliminate the radioactive nucleotides not incorporated in the probe. The column is already provided with a buffer that needs to be re-suspended by inverting several times. Later the column was spun 2 min. at 8 000 rpm. The buffer was discarded, the radio-labelled probe was then added in the middle of the column and later spun for 2 min. at 8000 rpm. After this step, the radiolabelled probe is now pure and is resuspended in 10 mM Tris buffer pH 7.4, ready to be used for Southern blot analysis.

2.2.12. Southern blot analysis

After DNA electrophoresis, the DNA was transferred following the standard protocol described by Sambrook et al., 1989, especially for pulsed field gels containing long DNA (Sambrook et al., 1989). DNA transfer was also performed by suction in a transfer machine (HYBAID) using 0.4 M NaOH buffer for 2 hrs. The Hybond membrane with the transferred DNA was dried and stored at room temperature in filter paper.

Later, in order to start the hybridization with the radio-labelled probe (2.2.10), 10 mg/ml sonicated salmon sperm DNA was denatured at 95°C for 5 min and spun for few seconds. The membrane was then neutralized in 2X SSC, 0.2M Tris-HCL pH 7.4 for 10 seconds, and baked at 80°C for 30 min. Then 150 µl of denaturated salmon sperm DNA was added to 15 ml of MIB buffer (1.5X SSPE, 10% PEG-8000, 70% SDS). This solution was called pre-hydrisation buffer (1.5X SSPE, 10% PEG-8000, 70% SDS, 100 µg/ml denatured sonicated salmon sperm DNA). The DNA membrane was put into a hybridisation flask and the pre-hybridisation buffer was added to it avoiding the formation of bubbles. Later, everything was incubated at 65°C for 1 hr with slow rotation.

The radioactive probe was denatured at 95°C for 5 min, put on ice for 2 min, spun down and added to the pre-hybridization buffer already in the tube. This is now hybridisation buffer. The filter was incubated at 65°C with slow rotation overnight.

Next day, the membrane was washed to remove the radiolabelled buffer first with wash buffer 1 (2X SSC, 0.1% SDS) at room temperature for 10 min shaking a bit. Then wash

buffer 2 (0.1X SSC, 0.1 % SDS) was added at 65°C. The filter was shaken for 5 min to 30 min, depending on the intensity of the radioactive signal. The membrane was then covered with Saran wrap making sure there were no bubbles and taped to a thin flat surface. Then in the dark room, the radiolabelled membrane was exposed to a Hyperfilm-MP (Amersham) inside a cassette with intensifying screens and kept overnight at -70°C. Next day the Hyperfilm was developed.

2.2.13. Preparation of digoxygenin (DIG)-labeled antisense RNA probes for *in situ* hybridisation.

First, 10 μg/μl of plasmid DNA template, containing either the 3'UTR-human *PDGFRA*, the ECD-mouse *Pdgfra*, *LacZ* gene or mouse *Olig2* gene sequence (Table 2.1), was linearised with the appropriate restriction enzyme in 100μl restriction reaction (total volume) and incubated at 37°C overnight (2.2.8). If considered necessary to inactivate the restriction reaction, 5μl of 10%SDS and 1μl of proteinase K (20 mg/ml) was added and the incubation continued at 37°C for 15 min. Phenol/chloroform extraction (2.2.3.) must follow this inactivation step to eliminate all the degraded proteins and the proteinase K from the DNA. The DNA was then precipitated with ethanol and resuspended in 50 μl of DEPC-treated distilled water or Baxter water. 1μg of this linearised DNA plasmid template was added to a 25 μl *in vitro* transcription reaction [5μl of 5X transcription buffer, 7.5 μl 100mM DTT, 2.5μl 10X DIG RNA labelling mix (10 mM of ATP, GTP,

CTP, 6.5mM UTP and 3.5 mM DIG-11-UTP, pH 7.5; Boehringer Mannheim), 1 µl of RNAsin (ribonuclease inhibitor, 20 units; Promega), 2 µl RNA polymerase (T3, T7 or SP6, 20 units, Promega), and 5.5µl DEPC-treated water]. Incubation at 37°C for 2 hrs was enough to synthesise the DIG-mRNA probe. The reaction was then set on ice, stopped with 1 µl of 0.5M EDTA pH 8 (DEPC-treated), made up to 100µl of total volume with DEPC water, quickly aliquotted and frozen at -80°C. 2µl of the probes were run on a 1% agarose gel with DEPC-1X TAE buffer.

Probe	**Insert and vector**	**Antisense transcript**
Mouse *Pdgfra*	1.6 kb EcoRI extracellular domain (ECD) cDNA fragment in pBluescript KS. From Chiayeng Wang and Nigel Pringle.	Cut with Hind III, transcribe with T7 RNA polymerase (T7 pol).
Human *PDGFRA*	1.7 kb EcoRI-NotI EST 3'UTR cDNA fragment in pT7T3D-Pac (Pharmacia).	Cut with EcoRI, transcribe with T3 RNA polymerase to generate 1.5 kb probe

LacZ	3.7 kb BamHI fragment of the β-galactosidase gene, cutting out its own ATG, from the pROSA-βgal plasmid (donated by P.Soriano (Soriano et al., 1991), and ligated in pBluescript-SK vector. This plasmid was named LacZpBlueSK.	Cut with XhoI, transcribe with T3 RNA polymerase for antisense probe. Cut with SacII, transcribe with T7 RNA polymerase for the negative control sense probe.

Table 2.1. Plasmid details and instructions for the synthesis of DIG-antisense mRNA probes for *in situ* hybridization.

2.2.14. Polymerase Chain Reaction (PCR)

Common PCR reactions for artificial chromosome characterization, transgene characterization and genotyping were done using the Taq polymerase (Promega), dNTPs (Bioline) and primers synthesized by the MWG-Biotech company. In case of requiring long PCR products or high quality amplified sequences for cloning, I used the expand high-fidelity PCR System (ROCHE). All the primer sequences, including the primers used for the end-rescue of YAC29E11 insert by inverse PCR, are listed in table 2.2. Furthermore, the conditions for each PCR reactions are specified in table 2.3.

Sequence	Primer sequence		Size Product
Human PDGFRA			
Promoter	5'-TGA AGG GAC ATG TGG AAG TAA TAG-'3	Forward	1095 bp
	5'- TTG AAA AGA AGA AAA GGA ACA TA-'3	Reverse	
Exon 1	5'-CTG GAC ACT GGG AGA TTC GGA G-'3		280 bp
	5'- CGA TGT TAT TCC GCA ATG AAT G-'3		
Exon 2	5'-CAG AAG GTT TTG GCT TCA GG-'3		508 bp
	5'-AAC TGC CAC TGG AGA GCA TT-'3		
Exon13&14	5'-GGA AGG TGG TTG AAG GAA CA-'3		
	5'-CAG GCT CCC AGC AAG TTT AC-'3		172 bp +intron
Exon 17-18	5'-CTA TCA AGT TGC CCG AGG AAT-'3		257bp +intron
	5'-GAC GTT GCG AGC AGC CAG AT-'3		
Exon 20-21	5'-TGT TCT ATC ATG CCA AGT GTT TCA-'3		389 bp
	5'-TTT ATA TTG TCC AGG CAG CAG ATT-'3		
Exon 23	5'-GGC CAC AGT CCT AGG TCT AGT-'3		226 bp
	5'-CTC GAA TCC GCC AGT TAC AGG-'3		
3'UTR	5'-GTG ATG TCC TTA AAA TGT GGT-'3		347 bp
	5'-GTA ATA CAT TTT GTA TTG GTA G-'3		
Upstream non-coding sequence			
30 kb	5'- CCC GGG TTG TCC ATC AGA GTT AT-'3		455 bp
	5'-TGG GAG GTG GAG GTT ACA GTG AG-'3		
60 kb	5'-GTC TCC ACT CTG CCT CTG CTA ACG-'3		662 bp
	5'-CTC CCA CCT CAT CCT CCC AAG TA -'3		
Downstream non-coding sequence			
9 kb	5'-CAG GGA GGA CAC AAC CAA AAC AAC-'3		474 bp
	5'-AGA AGT GGT GAG AGT GGG CAT CCT-'3		
10 kb	5'-GCC CAC TCT CAC CAC TTC TTT TCA-'3		485 bp
	5'-AGG TTG CTT TTG GTG GTA TGG TCA-'3		
20 kb	5'-GAT CTG GCC TTC GTC ACC TCT T-'3		574 bp
	5'-AGC TCC CCC ATA AAT CCT TCT GT-'3		
47 kb	5'-CTG CAG GAC CGG GGA GAC ACT-'3		487 bp
	5'-GGG CGA GAA GGG CTT GAA AAC T-'3		
70 kb	5'-GGG AGG CCG AGG CAG GTG TAT-'3		432 bp
	5'-TGA TCC CAG TAA GCA GAA GTG AGG -'3		
105 kb	5'-CTA ATC ACC AGG GGC ACA GTC TCT-'3		570 bp
	5'-CTC CTT ATC CAT CCC GTC TCA TCC-'3		

LacZ	5'-CCG AAA CTG TGG AGC GCC GAA ATC-'3 5'- AAA TCA CCG CCG TAA GCC GAC CAC -'3	884 bp
YAC left arm	5'-CAC CCG TTC TCG GAG CAC TGT CCG ACC GC -'3 5'-CCT TAA ACC AAC TTG GCT ACC GAG A -'3	250bp
YAC right arm	5'-ATA TAG GCG CCA GCA ACC GCA CCT GTG GCG-'3 5'-GTA ATC TTG AGA TCG GGC GTT CGA -'3	180bp
YAC left end	5'-TTA AAA CCA GAC TCA AAT C-'3 5'-ACA CTA GGT AAC CCA GAA CT-'3	108 bp

YAC INSERT ENDS RESCUE EXPERIMENT

Left end	5'-GTA GCC AAG TTG GTT TAA GG-'3 5'-ATA CAA TTG AAA AAG AGA TTC C-'3 5'-GGA CGG GTG TGG TCG CCA TGA TCG CG-'3
Right end	5'-AGT CGA ACG CCC GAT CTC AA-'3 5'-GAC TTG CAA GTT GAA ATA TTT CTT TCA AGC-'3 5'-AAG AGT CGC ATA AGG GAG AG-'3

Table 2.2. Primer sequences and length of PCR products.

	Human PDGFRA promoter	Exon 1	Exon 2	Exon13&14	Exon 17-18
PCR conditions:	**I**	**II**	**III**	**IV**	**I**
Genomic DNA	2 µl	2	0.5	0.5	2
10X PCR Buffer	2.5	2.5	2.5	2.5	2.5
25mM $MgCl_2$	2.5	2	2.5	1.5	2.5
20mM dNTPs	0.25	0.25	0.3	0.25	0.25
Taq polymerase	0.2	0.2	0.2	0.2	0.2
Primers (100pmol/µl):					
Up	0.1	0.1	0.1	0.1	0.1
Low	0.1	0.1	0.1	0.1	0.1
dH_20	17.35	18.45	18.8	19.85	17.35

		A			
PCR program:	94°C 10'	94°C 10'	94°C 10'	94°C 10'	94°C 10'
#cycles	35X 94°C 45" 48°C 45" 72°C 1'	34 X 94°C 30" 53°C 45" 72°C 1'	35 X 94°C 1' 65°C 1' 72°C 1'	33 X 94°C 1' 60°C 1' 72°C 1'	35 X 94°C 1' 57°C 45" 72°C 1'
	72°C 10'	72°C 10'	72°C 5'	72°C 5'	72°C 10'
PCR product length:	1095 bp	280 bp	508 bp	172 bp +intron	257bp +intron

	Upstream sequence 30 kb and 60 kb	Downstream non-coding sequence 9kb, 10kb and 20kb	47kb, 70 kb and 105kb	LacZ
PCR conditions:	I	I	I	V
Genomic DNA	2	2	2	2
10X PCR Buffer	2.5	2.5	2.5	2.5
25mM $MgCl_2$	2.5	2.5	2.5	2.5
20mM dNTPs	0.25	0.25	0.25	0.25
Taq polymerase	0.2	0.2	0.2	0.2
Primers (100pmol/µl):				
Up	0.1	0.1	0.1	0.5
Low	0.1	0.1	0.1	0.5
dH_20	17.35	17.35	17.35	16.5
PCR program:	94°C 10'	94°C 10'	94°C 10'	94°C 5'
#cycles	35 X 94°C 1' 55°C 45" 72°C 1'	35 X 94°C 1' 62°C 45" 72°C 1'	35 X 94°C 1' 60°C 45" 72°C 1'	35 X 94°C 45" 55°C 45" 72°C 1'
	72°C 10'	72°C 10'	72°C 10'	72°C 10'
PCR product length:	30kb upst: 455 bp 60kb upst: 662 bp	9kb downst: 474 bp 10kb downst: 485 bp 20kb downst: 574 bp	47kb downst: 487 bp 70kb downst: 432 bp 105kb downst: 570 bp	884 bp

	Exon 20-21	Exon 23	Exon23 3'UTR	YACleft and right arm	YACLeftend
PCR conditions:	I	II	II	II	VI
Genomic DNA	2	2	2	2	5
10X PCR Buffer	2.5	2.5	2.5	2.5	2.5
25mM $MgCl_2$	2.5	2	2	2	2.5
20mM dNTPs	0.25	0.25	0.25	0.25	0.25
Taq polymerase	0.2	0.2	0.2	0.2	0.2
Primers (100pmol/µl):					
Up	0.1	0.1	0.1	0.1	0.05
Low	0.1	0.1	0.1	0.1	0.05
dH_20	17.35	18.45	18.45	18.45	14.45
PCR program:		A	A		
	94°C 10'	94°C 10'	94°C 10'	94°C 3'	94°C 5'
#cycles	35 X	34 X	34 X	35 X	35 X
	94°C 1'	94°C 30"	94°C 30"	92°C 30"	94°C 45"
	50°C 45"	53°C 45"	53°C 45"	62°C 45"	38°C 45"
	72°C 1'	72°C 1'	72°C 1'	72°C 45"	72°C 1'
	72°C 10'	72°C 10'	72°C 10'	72°C 10'	72°C 10'
PCR product length:	389 bp	226 bp	347 bp	YACLeftarm: 250bp YACrightarm: 180bp	108 bp

Table 2.3. PCR reaction conditions and thermocycler programs.

2.3 Cell Biology

2.3.1. Preparation of mice embryos for cryostat tissue sections.

2.3.1.1. For *in situ* hybridization

Mouse embryos were extracted, rinsed with 1X PBS, and fixed with 4% PFA overnight. Embryos more than 14.5 days old were cut in half or three sections to allow the complete absorption of the fixative in all the tissues. The fixed embryos were then embedded in 20% sucrose until they sunk. They were dried with a tissue and embedded in blocks of Tissue Tek O.C.T. compound (SAKURA) at the appropriate orientation. All the bubbles were removed, and the blocks were put on a 10 cm diameter metallic tray containing 30 ml of methanol, then the tray was quickly filled with dry ice and the blocks were frozen in seconds. All the blocks were labelled and stored at -70°C.

2.3.1.2. For β-Gal staining in cryostat sections

After the mouse embryos were extracted, and rinsed with 1X PBS, they were then fixed with X-Gal Fixative (1% Formaldehyde, 0.2% Glutaraldehyde (GDA), 2mM $MgCl_2$, 5mM EGTA and 1X PBS) by 5 to 60 min, according to stage/size, at 4°C. E11 embryos were fixed during 1 hr, E12.5 and E13.5 during 2 to 3 hrs, and half P8 Brain around 3 hrs. Embryos more than 14.5 days old were cut in half or three sections to allow the complete absorption of the fixative in all the tissues. The fixed embryos were then put in 20% sucrose until they sunk, then dried with a tissue and embedded in blocks of Tissue Tek O.C.T. compound (SAKURA), where all the bubbles were removed. In order to freeze the embryos as soon as possible, the blocks were put on a 10 cm diameter metallic tray

containing 30 ml of methanol, then the tray was quickly filled with dry ice and the blocks were frozen in seconds. All the blocks were labelled and stored at -70°C.

Another method used for this staining, was by rinsing the embryos only with 1X PBS, drying them with a tissue and then immediately freezing them quickly in tissue tek, without any fixation or sucrose steps. In this case, only 20 μm cryostat tissue sections were cut carefully, let dry for one hour in a fume hood, and then fixed with the X-Gal Fixative for 20 min at room temperature. The quality of the sections is not as good as using the previous method, but the X-Gal staining works perfectly and the advantage is that other cryostat sections of this embryo could be fixed differently and then used for immunohistochemistry or even *in situ* hybridization.

2.3.2. Treatment with Proteinase K (PK) for *in situ* hybridisation.

The preparation of the tissue was done as described in 2.3.1.1. Frozen cryostat tissue sections were cut and put on slides treated with Vector Bond. The sections were let dried at least 45 minutes. Only for proteinase K treatment, the sections were rehydrated twice with 1X PBS by one minute each in a humidified chamber. PK was diluted in PK buffer (10X stock: 500mM Tris-HCl pH 7.5, 50 mM EDTA) to 20 μg/ml for adult tissue or 10μg/ml for embryonic tissue. The PK solution was added to the sections and incubated at room temperature by 7.5 min. After this PK digestion, the sections were immediately washed with 0.2 % Glycine (in DEPC-treated water) for 30 seconds to stop the reaction, and later washed with 1X PBS twice by 30 seconds each. A step of fixation with 4%

paraformaldehyde by 20 min was then required before the dehydration through alcohols (from 30% to 100%). The sections were then dried out again.

2.3.3. Immunohistochemistry

Frozen tissue sections were cut, preferably 15 μm thick, with the cryostat, put on coated slides with Vector Bond, and let dry at least for 45 min in a fume hood. The type of fixative and time of fixation was determined depending on the specific antibodies, normally light fixation with 4% PFA before tissue embedding was enough to maintain the proteins.

Once the tissue sections are completely dried, the slides were covered with blocking solution (10% (v/v) goat serum) and incubated in a humidified chamber at room temperature for 2 hours. Later, the first antibody was diluted in 1X PBS with 0.1% (v/v) triton and 10% (v/v) goat serum to a concentration around 1:100 and 1:1000, and added 200 μl per slide. The slides were incubated in a humidified chamber at 4°C, overnight. Next day, the first antibody was washed with 1X PBS, 0.1 % (v/v) Triton buffer three times 5 minutes each. The second antibody normally fluorescent, either with fluorescein (Green) or rodamine (Red), was diluted around 1:250 or 1:2000, and incubated at room temperature during one hour. The second antibody is rinsed and washed again with 1X PBS, 0.1% Triton buffer three times 5 minutes each. The slides were mount with CITIFLUOR (Glycerol/PBS solution).

2.3.4. Single *in situ* hybridisation using digoxigenin (DIG)-labelled RNA probes and NBT/BCIP staining.

While the transgenic founders were genetically characterised, I also started the analysis of the transgene expression by β-Gal staining (sections 2.3.1.2 and 2.3.6) and/or *in situ* hybridization (2.3.1.1 and 2.3.4). In order to get a Dig-LacZ probe for *in situ* hybridization, I excised a 3.7 kb BamHI fragment of the *β-galactosidase* gene, cutting out its own ATG, from the *pROSA-β gal* plasmid (donated by P. Soriano (Soriano et al., 1991), and ligated it into the BamHI linearised pBluescript SK vector (Table 2.1) (procedures described in 2.1.2, 2.1.3, 2.2.8, 2.2.3, 2.2.4, and 2.2.1). This plasmid was named LacZpBlueSK. The Dig-labelled *LacZ* mRNA probe was then synthesised (2.2.13) and first tested in the *LacZ* positive Rosa26 transgenic line (data not shown) before use for analysis of the expression studies.

The production of other specific DIG-RNA probes is described in 2.2.13 (check table 2.1). Dilutions of the antisense DIG-RNA probe were done in 1 ml of hybridisation buffer (1X salts, 50% deionised formamide, 10% (w/v) dextran sulphate, 1X Denhardt's solution and 1 mg/ml yeast RNA. 10X salts stock: 2M NaCl, 100mM Tris-HCl pH 7.2, 50mM NaH_2PO_4, 50mM Na_2HPO_4 and 50mM EDTA). The extracellular domain mouse *PDGFRA* probe was diluted to 1:500; the 3'UTR human *PDGFRA* probe to 1:1000 and the mouse *Olig2* probe to 1:1000, then vortex, centrifuge briefly, and denatured at 75°C for 5 min. Each slide was covered with 200 µl of hybridization buffer with the specific probe and coverslipped avoiding any bubbles. The slides were incubated at 65°C in a sealed perspex box with two sheets of Whatman paper wetted in 1X SSC, 50%

formamide, overnight. Next day, the slides were washed with wash buffer (1X SSC, 50% formamide, 0.1% Tween-20) pre-warmed to 65 °C by 30 min in a coplin jar. The next two washes were made with 5X MABT (100mM maleic acid pH 7.5, 150 mM NaCl, 0.1% Tween-20) pre-warmed to 65°C by 30 min each, then two more washes with 1X MABT at room temperature by 30 min. After the washes, 300 µl per slide of blocking solution (MABT containing 2% blocking reagent (Boehringer Manheim), 10% heat-inactivated sheep serum) was added and incubated for one hour at room temperature. In order to detect and labelled the DIG-RNA probe, the antibody anti-Digoxigenin-alkaline phosphatase (Fab fragments, Boehringer Mannheim) was diluted to 1:1000 in blocking solution, added to each slide and incubated overnight at 4°C.

The slides were washed twice with MABT 1X by 5 minutes each, then three times with prestaining buffer (100mM Tris-HCl pH 9, 100mM NaCl, 5 mM $MgCl_2$) by 5 minutes each. After washing, 40 ml of staining buffer (10% polyvinyl alcohol (Mw 70,000-100 000, BDH), diluted 1:1 with prestaining buffer, plus 87 µl of 100mg/ml nitroblue tetrazolium salt (NBT dissolved in 70% DMF, Boehringer Mannheim), 67 µl of 50mg/ml 5-bromo-4-chloro-3 indolyl-phosphate (BCIP, Boehringer Mannheim), and 1.9 ml of 1M MgCl) was added in a coplin jar then incubated at 37°C overnight in dark. The staining was checked on a microscope, without letting the slides dry. Once the staining was strong and clear enough, the slides were washed twice with tap water for 1 min, and then dehydrated with a series of ethanol (from 30%, 60%, 80%, 96% to 100% 1 min each, and then twice with Xylol first 5 min and then at least for 30 min. The slides were then mounted with Gurr and dried.

2.3.5. Double fluorescent *in situ* hybridisation

Two antisense mRNA probes, one DIG-labelled (2.2.13) and the other FITC labelled (using 3.5 mM FITC-12-UTP, Roche), were used in this double *in situ* hybridisation method. Both probes were diluted 1/1000 in hybridisation buffer (1X salts, 50% deionised formamide, 10% (w/v) dextran sulphate, 1X Denhardt's solution and 1 mg/ml yeast RNA. 10X salts stock: 2M NaCl, 100mM Tris-HCl pH 7.2, 50mM NaH_2PO_4, 50mM Na_2HPO_4 and 50mM EDTA), denatured at 75°C for 5 min, 150µl of this solution was added on each slide (Table 2.1). A coverslip was slowly laid on top of the slides to avoid evaporation, and they were let hybridise overnight at 65°C in a sealed container with Whatman filter paper soaked in 2X SSC plus 50% formamide. After hybridisation, the slides were immersed in 1X MABT (100mM maleic acid pH 7.5, 150mM NaCl, 0.1% (v/v) Tween-20) until the coverslips fell off and washed again three times for five minutes each in 1X MABT at 65°C before placing them in wash buffer (1X SSC, 50% formamide, 0.1% Tween-20) also at 65°C twice for 30min each. The slides were then incubated twice for 30 minutes each in 1X MABT at room temperature. Blocking solution (1X MABT containing 2% blocking reagent-Roche and 10% heat-inactivated sheep serum) was added to the slides followed by incubation for 1 hr at room temperature.

Different to the previously described single *in situ* hybridisation using an alkaline phosphatase (AP)-conjugated anti DIG antibody (2.3.4), this double *in situ* hybrisation uses horseradish peroxidase (HRP)-conjugated antibodies; either anti-DIG-POD (Roche, 1-207-733) or anti-FITC-POD (Roche, 1-426-346). Depending on which was the first probe, either DIG or FITC, the appropriate antibody was diluted (1/200 for anti-DIG-POD

or 1/1500 for anti-FITC-POD) in blocking buffer and incubated overnight at 4°C. The slides were later washed three times 10 min each, in 1X PBS containing 0.1% Triton X100.

In order to get a fluorescent signal, the tyramide signal amplification (TSA) plus fluorescent system was used, which catalyzes a fluorophore labelled tyramide amplification reagent with the HRP enzyme. The TSA Plus kit components (PerkinElmer Life Sciences INC, NEL741-NEL745B) were mixed just before staining, making a 1/50 dilution of the fluorescent tyramide (fluorescein, rhodamine, or cyanine) with the amplification buffer provided. 100 µl per slide of fluorescent mix solution were added and a coverslip was placed over each slide. Incubation was for 10 min at room temperature. Three washes five min each with 1X PBS containing 0.1% Triton X 100 followed to remove the staining solution.

In this method, it was essential to inactivate the HRP enzyme after the first fluorescent staining reaction in order to avoid cross-developing reactions causing false positive co-localization of the probes analysed. The HRP enzyme was inactivated by incubated the slides in 3%H_2O_2 in 1X PBS for 30 min at room temperature in a coplin jar. After HRP inactivation, the slides were washed with 1X PBS with 0.1% Triton X 100 several times 15 min each to remove completely the H_2O_2, and then blocked again with blocking solution for 1 hr before adding the second HRP-antibody. Sections where either the second probe or second antibody was omitted were used as controls for specificity of the second layer reaction. The fluorescent staining reaction was done again using a different

fluorescent tyramide to the one used to develop the first probe, usually fluorescein and rhodamine were the most common combination.

2.3.6. β-Galactosidase staining.

2.3.6.1. Wholemount staining of embryos

Mouse embryos were extracted and rinsed in 1X PBS. They were then fixed with X-Gal Fixative (1% Formaldehyde, 0.2% Glutaraldehyde (GDA), 2mM $MgCl_2$, 5mM EGTA and 1X PBS) for 5 to 60 min, according to stage/size, at 4°C. E11 embryos were fixed for 1 hr, E12.5 and E13.5 during 2 to 3 hrs, and Half P8 Brain around 3 hrs. After fixation, they were washed with wash buffer (0.1 M phosphate buffer pH 7.3, 2mM $MgCl_2$, 0.1% sodium deoxycholate, 0.02% NP40) three times by twenty minutes each at room temperature. Later, the embryos were incubated in X-Gal staining solution (1ml of 50 mg/ml X-Gal (5-bromo-4-chloro-3-indolyl-B-D galactosidase, Invitrogen), 82 mg K_3Fe (CN_6) (Sigma), 105 mg K_4Fe (CN_6) $3H_20$ (Sigma), diluted in 49ml wash buffer) overnight or longer at 30°C in the dark. The embryos were covered completely with the staining solution and without rotation. After staining, they were rinsed again with wash buffer twice for 5 min each. The photographs were taken the same day because the staining starts diffusing after a few weeks, even though the stained embryos were stored in X-Gal fixative at 4°C. Alternatively, they were fixed again and incubated in 4% sucrose at room temperature until they sank. They were then embedded in blocks of Tissue Tek O.C.T. compound (SAKURA), all the bubbles removed, and frozen on a metallic tray with methanol on dry ice, labelled, and stored at -70°C.

2.3.6.2. Cryostat tissue sections

Frozen sections were cut 20μ thick, which was appropriate for this kind of staining. Always, sections were cut from a negative control and were processed with the experimental ones to check any endogenous β-Gal activity. The sections were let dry for one hour at least. If the sections were not going to be used immediately they were stored at -70°C. The sections were covered with 300 μl of X-Gal fixative (1% (v/v) formaldehyde, 0.2% (v/v) Glutaraldehyde, 2mM $MgCl_2$, 5mM EGTA, 1X PBS) in a humidified chamber, and then incubated for 20 min at room temperature. Preferably, the fixative stock was protected from light and stored at 4°C. The slides were then washed three times for ten minutes each with wash buffer (0.1 M Phosphate buffer pH 7.3, 2mM $MgCl_2$, 0.1% Sodium deoxycholate, 0.02% NP40). Staining buffer was then protected from light and incubated at 30°C overnight. The staining buffer stock (1ml of 50 mg/ml X-Gal (5-bromo-4-chloro-3-indolyl-B-D galactosidase, Invitrogen), 82 mg K_3Fe (CN_6) (Sigma), 105 mg K_4Fe (CN_6) $3H_20$ (Sigma), diluted in 49 ml wash buffer) is mixed and protected from light, there was no need to filter it and it could be stored for 5 days at 4°C. This incubation was sometimes longer than overnight if the staining was still weak, it never acquired background. After staining, the slides were washed three times for 5 minutes each in 1X PBS, post-fix in X-Gal fixative by 10 min at room temperature, washed again with 1X PBS three times by 5 min each, dehydrate through alcohols and xylene (30% ETOH by 15'', 60% ETOH by 15'', 80% ETOH by 30'', 95% ETOH by 1', 100% ETOH by 1', Xylene by 1'), and then mounted in permount (Xam, Gurr). The β-Gal staining on sections started to diffuse after two or three months.

2.4. Yeast Artificial Chromosome (YAC) production and purification.

2.4.1. Yeast strains and storage

The *Saccharomyces cerevisiae* strains CEPH 29E11 (Spritz et al., 1994), YLBW3 (Hamer et al., 1995) and three new strains containing YACs modified from the original contained in the CEPH 29E11 were used in this research project. These three strains were cloned by Els Wessels and Dr. Gijs Afink (Wessels E., 2001). The human *PDGFRA* gene was fragmented and the *nls-Lac Z* reporter gene was introduced after exon1 in these YAC constructs. Their last step of cloning was to transfect again these YACs into the YLBW3 host strain. These constructs and their corresponding name are shown in figure 3.1.

Yeast strains were grown at 30°C in their respective liquid selective media (Table 2.4) or on agar plates containing liquid selective media with 25g/L bacto-agar. Ampicillin was added to the liquid or molten solid media (after cooling to 55°C) at a final concentration of 50-100 μg/ml (100 mg/ml stock in water, 0.22 μm filter-sterilised and stored in aliquots at -20°C). Liquid cultures were continually agitated in a rotating environmental shaker at 250 rpm at 30°C.

For short term storage of yeast strains and clones, they were stored on agar plates at 4°C for two or three weeks. For long term storage, 75% glycerol stock was added to the pellet of overnight yeast liquid cultures to a final concentration of 15% glycerol. This mixture

was stored in 0.5-1 ml aliquots at -20°C or 80°C. This long term stocks could keep viable for at least two years.

Yeast strain	**Culture medium**	**Reactives**
YAC 29E11	Selective Yeast medium (1)	0.72 g –Trp/-Ura DO supplement (Clontech), 26.7 g SD Minimal Base pH 5.8 (Clontech), and 20 g of D(+)-Glucose/ 1lt .
YAC 29E11LacZexon 2-F	Selective Yeast medium (2)	0.62 g –Leu/-Trp/-His DO supplement (Clontech), 26.7 g SD Minimal Base pH 5.8 (Clontech), 20g of D(+)-Glucose and 10ml of 30mg/ml L-leucine/ 1lt .
YAC 29E11LacZexon 23-F	Selective Yeast medium (3)	0.62 g –Leu/-Trp/-His DO supplement (Clontech), 26.7 g SD Minimal Base pH 5.8 (Clontech), 20g of D(+)-Glucose and 10ml of 30mg/ml L-leucine/ 1lt .
YLBW3	YPD (for routine growth)	10g of 1% Bacto-yeast extract, 20g of 2% Bacto-peptone, 20g of 2% Dextrose, and 20g of 2% Bacto-agar / 1 lt.

Table 2.4. Yeast strains and their selective culture media. All the yeast media were autoclaved at 121°C for 15 min, and stored at 4°C.

2.4.2. Preparation of low density plugs containing YAC DNA.

The YLBW3, CEPH 29E11 and CEPH 29E11 Lac Z strains were grown in selective media plates at 30°C for two or three days. Medium and small yeast colonies were picked by yellow tips to grow in 5 ml of its respective selective YCD medium (Table 2.4) at 30°C, 250 rpm overnight. This 5 ml culture was transferred to 50 ml YCD medium to grow at 30°C 250 rpm for 2 days approximately.

The yeast culture was then spun at 3000 rpm for 5 minutes. The pellet was washed in 15 ml of 50mM EDTA (pH 8.0) twice, resuspended in 1.5 ml of 50 mM EDTA, mixed gently with cut off blue tips with 0.5ml SCE (1 M sorbitol, 0.1 M sodium citrate, 60 mM EDTA), 2.5 µl β-mercaptoethanol and 1 mg Zymolyase-100T (ICN Biomedicals) and 2.5 ml of 1% low melting point (LMP) agarose (NUSIEVE agarose) dissolve in 0.125 M EDTA. 100 µl aliquots of the mixture were pipetted using a cut off yellow tip into plug formers kept on ice, and left for 10 minutes to allow the agarose to set and to form plugs. The plugs were then transferred into 25 ml of 0.45 M EDTA, 10 mM Tris-HCl (pH 8.0) and 7.5% β-mercaptoethanol, incubated at 37°C overnight, removed from the above solution and replaced with 0.5M EDTA, 1% N-lauroylsarcosirate and 1 mg/ml Proteinase K, continually incubated at 50°C overnight and repeated it once, finally stored at 4°C in 0.5M EDTA, 1% N-lauroylsarcosirate and 10 mM Tris-HCl (pH 9.5).

These plugs were used for Southern blots and as markers for YAC DNA purification.

2.4.3. Preparation of high density plugs containing YAC DNA.

A single yeast colony was picked and grown in 5 ml of its respective selective media (Table 2.4) at 30°C, 250 rpm overnight. This small culture was then used to inoculate 500 ml of selective media in a 2 L flask, which was incubated at 30°C, 250 rpm for about two or three days depending on the yeast strain. The yeast density was checked by counting yeast cells using a Haemocytometer, until the culture reached a concentration of 2-4 x 10^7 cells per ml. The yeast culture was spun at 4°C, 4000 rpm for 5 minutes. The pellet was resuspended in 50 ml SE buffer (1M sorbitol, 20mM EDTA pH 8.0) and then washed twice in SE buffer. After the last wash, the supernatant was carefully removed with a paper towel and the pellet was resuspended in 200 µl of SE buffer with a cut off yellow tip. 0.5 ml aliquots of cell suspension were transferred to 1.5 ml Eppendorf tubes and incubated at 37°C. A 1% low melting point agarose (NUSIEVE agarose) was dissolved in SE buffer containing 14 mM β-mercaptoethanol and incubated at 42°C. Just before use, 10 mg of Zymolyase-100 T (ICN Biomedicals) was dissolved in 2 ml of the low melting point agarose solution. 0.5 ml of this solution was transferred into 0.5 ml yeast cell suspension and mixed thoroughly by pipetting up and down using a cut off blue tip. The mixture was then incubated at 42°C at all times preventing the agarose for setting. 100 µl of this mixture was poured into a hole of the plug former and kept on ice. By 10 minutes, the plugs set.

One by one, the plugs were carefully transferred to SE buffer containing 14 mM β-mercaptoethanol and 1 mg/ml Zymolase, in a volume of at least 0.5 ml per plug, and incubated at 37°C overnight. Next day, the buffer was replaced with a 0.2 M EDTA pH 8.0, 0.1 M Tris-HCl pH 8.0, 0.5 M NaCl, 0.1% SDS, 0.5M β-mercaptoethanol and 1 mg/ml proteinase K buffer. After an incubation at 37°C overnight , the plugs were washed in TE buffer pH 8.0 until no more bubbles could be seen and then stored in 0.5 M EDTA at 4°C until use.

2.4.4. DNA isolation from yeast colonies on a plate.

A sample from one of the yeast colonies was taken and added to 1.5 ml of 1:20 diluted β-mercaptoethanol, then vortexed briefly and boiled for 5 minutes. The sample was then incubated for 1 minute on ice and spun at 13 000 rpm by 2 minutes. The supernatant containing genomic yeast DNA was recovered and stored at -20°C. 25 µl of this DNA solution was used for PCR reaction.

2.4.5. Isolation of intact YAC DNA for microinjection

At the beginning of this PhD project YAC purification protocols were already published (Schedl et al., 1996). However, I made some optimization during the course of this research to adjust to the particular yeast strains and fragmented constructs that we designed to study the regulation of the Human PDGFRα gene. Most of those optimization

steps were done in order to increase the YAC DNA concentration yield per purification procedure, reduce the YAC DNA shearing and the time of purification (2.4).

During this PhD project, I followed the general method described by (Schedl et al., 1996;Sun, 1999) to purified the YAC DNA. However, I optimized these procedures for the specific detection, characterization and purification of our YAC constructs; especially the pulsed field gel electrophoresis (PFGE) (2.4.5.1). In addition, I followed the method of inverse PCR (Silverman G A, 1996) to identify precisely the ends of the DNA insert in the YACs (2.5).

2.4.5.1. Pulsed Field Gel Electrophoresis (PFGE)

Multi-purpose agarose (molecular grade from BIOLINE) was used to prepare a 1% gel with 0.25X TAE buffer. The melted agarose was poured in the CHEF-DR II casting stand with a comb with five wells sealed with tape (BIO-RAD). The low and high density plugs (2.4.3 and 2.4.2) were then washed with TE buffer (pH 8.0) on a shaking platform at 4°C, at least four times for 15 min each. Another 2 litres of 0.25X TAE buffer were prepared with ultra-pure water and were chilled on ice. The plugs were loaded next to one another into the preparative lane, including any marker plug (e.g. PFGE yeast marker, W3, 29E11/W3) as positive or negative controls to localise the YAC DNA. The plugs were sealed with 1 % agarose dissolved in 0.25X TAE. The cool 2 litres of 0.25X TAE buffer were poured into the electrophoresis chamber of the CHEF-DR II Pulsed field electrophoresis system (BIO-RAD). Fresh and cool 0.25X TAE buffer was always used to run a pulsed field gel, because it assured that the DNA ran perfectly and the

electrophoresis chamber would not overheat. The pump was turned on and it was checked that the tube did not contain bubbles. The correct PFGE program was set depending on the size of the YAC DNA to be purified (Table 2.5).

Program	**Y29E11 and Y29E11LacZ**	**Y29E11LacZ exon23-F**	**Y29E11LacZ exon 2-F**
Voltage	6 V / cm	6 V/ cm	6 V / cm
Run time	23 hr	36 hr	20 hr
Initial Switch time	25 sec	1 sec	1 sec
Final Switch time	45 sec	12 sec	12 sec
Temperature	4°C	4°C	4°C
DNA separation	610 – 225 kb	194 – 90 kb	194 -9.4 kb
Size of the YAC	453 kb	130 kb	90 kb
Marker	Yeast chromosomes	MidRange PFG	Low Range PFG
Photograph	Fig.2.1, A	Fig.2.1, B	Fig.2.1, A

Table 2.5. Pulsed field gel electrophoresis programs per yeast strain. Programs set in the CHEF-DR II Pulsed field electrophoresis system (BIO-RAD).

Once the electrophoresis finished, the marker lanes were cut off and stained in 0.25 X TAE buffer containing 0.5 μg/ml ethidium bromide (Fig 1, A). The position of the YAC DNA band and endogenous yeast chromosomes were localized in the stained markers lanes. With the help of a ruler, the gel was reassembled and three 5-well gel bands were excised, two were called marker 1 and 2, and the other was the YAC DNA band. These 5-well gel bands were not stained with ethidium bromide, specially the YAC DNA gel band.

2.4.5.2. Mini-electrophoresis

A 4% LMP agarose gel (NUSIEVE agarose) was prepared in 0.25 X TAE buffer (30 ml were enough). The agarose was melted and left to cool until it reached around 55-65°C. 600 ml of 0.25X TAE buffer were also prepared and cooled in the -20°C freezer. At the same time, the minigel electrophoresis chamber (H3 Mini Submarine Unit HOEFER) was cooled and put into the freezer for a few minutes. Then the YAC DNA band was put in the middle on a minigel tray and the two markers, one on each side of the YAC DNA band. An asymmetric order was set to identify easily the right position of the gel. Once cooled to around 60°C, this 4 % LMP agarose was poured in the minitray around the 1% PFGE gel cut lanes containing the YAC and markers DNA, and it was left to set in the fridge (Fig 1, B). Later, this gel was set into the cool mini-electrophoresis chamber, the cool 0.25XTAE buffer was poured in and the gel ran at 50 volts for 18 hrs. After electrophoresis, the two marker lanes were cut off and stained in 0.25 X TAE buffer with 0.5μg/ml EtBr for 20 min with gentle agitation. This allowed checking under a UV trasiluminator, to see if the marker DNA had already completely run into the 4% LMP agarose. If so, the gel was then reassembled and the position of the YAC DNA in the 4% LMP agarose gel was

determined, by comparison to the markers position, and then cut off in a 0.5-0.8 cm block. The YAC DNA was never stained with EtBr.

2.4.5.3. Agarose digestion

The 0.5 cm YAC DNA agarose block was equilibrated on a rocking platform for at least 1.5 hours with 12.5 ml of TENPA buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0, 100 mM NaCl) with 30 μM spermine and 70 μM spermidine added fresh to the buffer. The gel slice was dried with a clean tissue. The YAC DNA block was transferred into a 1.5 ml Eppendorf tube; it was then melted for 3 min at 68°C, centrifuged 10 seconds to bring down the molten agarose and incubated for additional 5 minutes at 68°C. After melting, the agarose was transferred quickly to the 42°C incubator and left for 5 min. It was very important to keep the DNA-Agarose solution completely melted at 42°C before adding the enzyme β-agarase I (New England BioLabs). The eppendorf tube was never kept out of the 42°C incubator for more than a few seconds, otherwise the solution would start to set and the enzyme would not work. Working in the 42°C bath, 2 units of β-agarase I (New England BioLabs, stock 1 unit per μl) was added per 100 μl of molten-agarose containing the YAC DNA, and was mixed very gently with a cut-tip using the micropipette. The vortex was never used to mix YAC DNA in liquid. The enzymatic reaction was incubated at 42°C for 3 hours. Later, this YAC DNA/Agarose solution was left at room temperature for 10 min and checked to see if the agarose was fully digested.

2.4.5.4. Dialysis

In order to prepare this YAC DNA for microinjection, first 5 ml of micro-injection buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA pH 8.0, 100 mM NaCl) was prepared with fresh 30 μM spermine and 70 μM spermidine, then mixed thoroughly and put in a small Petri dish (Fig 2.1). Three dialysis membranes (Millipore, pore size 0.025 μM) were left to float on the microinjection buffer. On each dialysis membrane was put 50 μl of the YAC DNA solution very carefully, and incubated at room temperature for 1 hour. The dialysed YAC DNA solution was recovered carefully from the top of the dialysis membrane and stored at 4°C.

2.4.5.5. Electrophoresis

The DNA concentration was checked by running 5 and 10 μl of purified YAC DNA on a thin 0.8 % gel with small wells and stained with 0.2μg/ml EtBr. λ DNA (New England Biolabs, or GIBCO BRL) was used as mass marker, 5 ng, 10 ng and 20 ng of λ DNA were loaded before the YAC DNA. The concentration of the YAC DNA was determined by visual comparison between the UV intensity of the DNA bands.

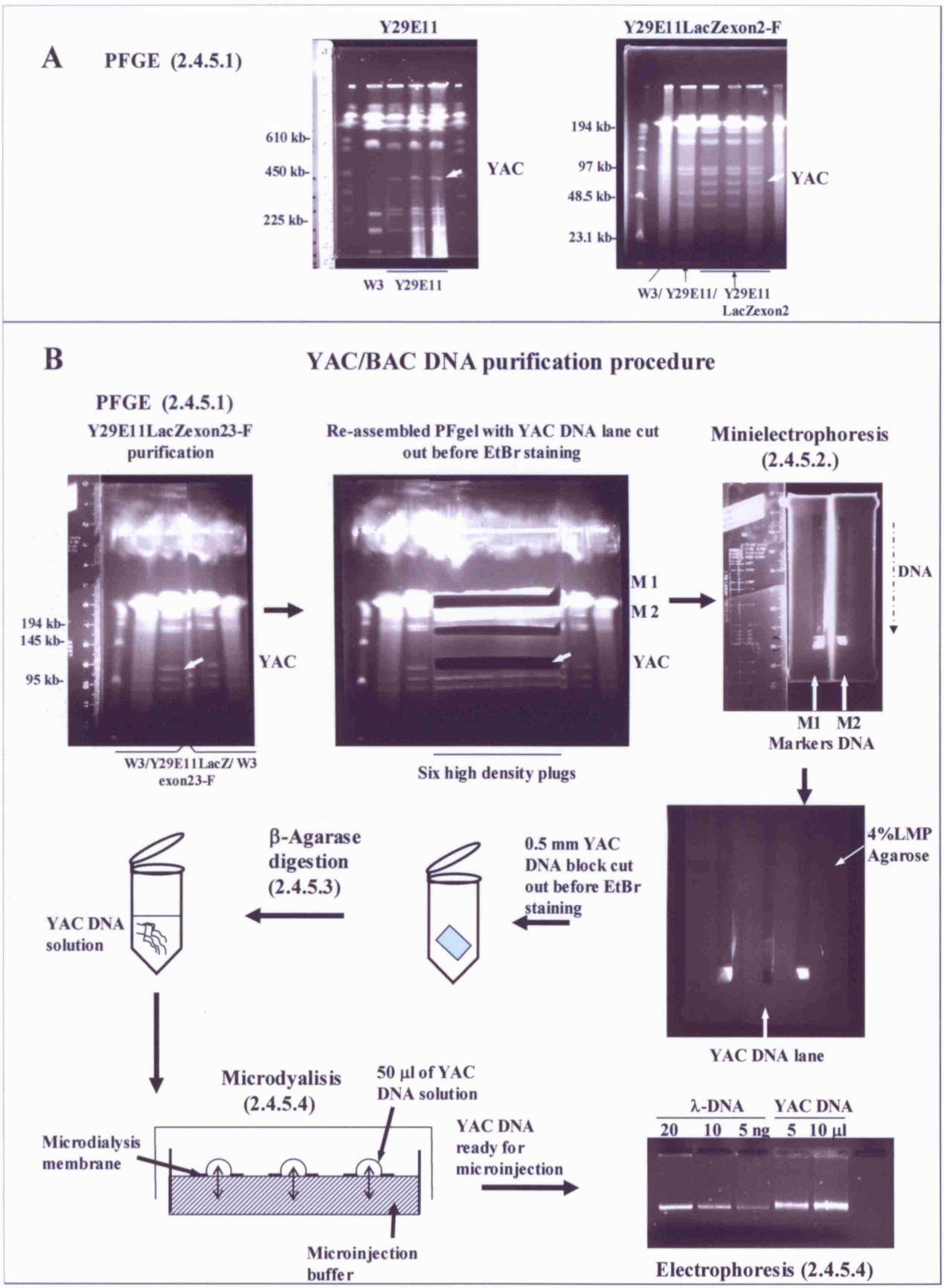

A
PFGE (2.4.5.1)
Y29E11
610 kb-
450 kb-
225 kb-
YAC
W3 Y29E11
Y29E11LacZexon2-F
194 kb-
97 kb-
48.5 kb-
23.1 kb-
YAC
W3/ Y29E11/ Y29E11 LacZexon2
B
YAC/BAC DNA purification procedure
PFGE (2.4.5.1)
Y29E11LacZexon23-F purification
194 kb-
145 kb-
95 kb-
YAC
W3/Y29E11LacZ/ W3 exon23-F
Re-assembled PFgel with YAC DNA lane cut out before EtBr staining
M 1
M 2
YAC
Six high density plugs
Minielectrophoresis (2.4.5.2.)
DNA
M1 M2
Markers DNA
4%LMP Agarose
YAC DNA lane
0.5 mm YAC DNA block cut out before EtBr staining
β-Agarase digestion (2.4.5.3)
YAC DNA solution
Microdyalisis (2.4.5.4)
50 µl of YAC DNA solution
Microdialysis membrane
YAC DNA ready for microinjection
Microinjection buffer
λ-DNA
YAC DNA
20 10 5 ng 5 10 µl
Electrophoresis (2.4.5.4)

Fig. 2.1. Representation of the general method for intact YAC DNA purification. A. PFGE of Y29E11 and Y29E11exon2-F for YAC DNA isolation without risk of contamination from endogenous yeast chromosomes. B. Purification of YAC DNA, using as an example the purification of the Y29E11LacZexon23-F DNA, starting from visualizing the three first and two last lanes of the PFgel used as markers to localize the YAC DNA. The location of the YAC in the unstained five sealed well-part of the PFgel was calculated with the use of a ruler. The same strategy was used to localize the YAC DNA after concentrating in a block by running a minigel. This pure-intact YAC DNA is very fragile in liquid form, so it was used for pronuclear microinjection in mice eggs the same day or next morning of the end of the isolation. Reference number to the text of the particular protocols is specified. The same method was used to isolate PAC and BAC intact DNA for microinjection.

2.5. Bacterial Artificial Chromosome (BAC) production and purification

2.5.1. Maxi BAC/PAC DNA preparation

Two methods could be used to purify either BACs or PACs. One method of maxi-purification of BAC/PAC DNA was the Nucleobond® BAC Maxi Kit from BD Biosciences (Cat # 635941), following the manufacturer protocol. The second method is an adaptation to the alkaline lysis method (2.2.1) for BAC/PAC DNA, and started by

inoculating 500μl of a single isolated BAC colony culture (Homo sapiens BAC RP11-626H4, from BACPAC Resource Center at C. H. O. R. I., California), or EL250 bacteria electroporated with the modified BAC (CE⁻human *PDGFRA* BAC-110kb), into 400ml of LB containing 15μg/ml chloramphenicol and incubated at 32°C, 16-24hrs, shaking at 250 rpm. The bacteria culture was divided between 8 50ml Falcon tubes then centrifuged at 3750 rpm for 15 minutes at 4°C. Supernatants were discarded and the pellets re-suspended in 8ml of solution P1 (50mM Tris-HCL, pH8.0, 10mM EDTA, pH 8.0, 100μg/ml RNase A) by pipetting up and down with cut-up 1ml tips. Gently pipetting and use of wider-cut tips were used during this method to avoid shearing of the BAC/PAC DNA. Then, 8ml of fresh P2 were added (0.2N NaOH, 1% SDS) and the tubes gently inverted several times, the suspension changed from turbid to almost translucent, and then incubated at room temperature for no longer than 5 minutes. 8ml of P3 solution (3M KOAc, pH 5.5, 11.5% w/v glacial acetic acid) were added and each tube immediately gently inverted. A thick white precipitate of protein and DNA formed. The tubes were then incubated on ice for one hour before centrifuging at 3750 rpm for 15 minutes at 4°C. Each supernatant was then transferred to a fresh falcon tube by passing through a double sheet of autoclaved mesh (filter cloth), and centrifuged again as before. Supernatants were transferred to a new falcon tube and 24ml of ice-cold isopropanol added, mixed by inverting a few times and left on ice for 5 minutes. The supernatants then were centrifuged at 3750 rpm for 15 minutes at 4°C. The pellets were re-suspended in 250μl of TE using wide cut tips to prevent DNA shearing, and the supernatants discarded. Cut-wider tips were used for all BAC/PAC DNA pipetting. All the pellets were pooled into 4 x 1.5ml eppendorf tubes. The DNA was spun for seconds at 13200 rpm, the supernatants transferred to new eppendorf tubes and added 500μl of phenol-chloroform, pH 7.9. The tubes were gently inverted 8

times and spun at 13200 rpm for 5 minutes. The supernatant were transfered to fresh tubes and the phenol-chloroform extraction repeated. Then, 500μl of chloroform:IAA was added to each supernatant, mixed carefully by inverting 8 times, spun again at 13200 rpm for 5 minutes, and pooled into one 2ml eppendorf tube. Again, this DNA was then split into 4 x 1.5ml eppendorf tubes and added 1ml ice cold 100% ethanol mixed again by inverting 4 times. To keep a stock, 2 tubes could be stored at -20°C. The rest of the tubes were spun at 13200 rpm for 10 minutes. The BAC DNA pellets were washed in 1ml room temperature 70% ethanol by inverting gently and spinning again. After aspirating all the ethanol, the pellets were air dried at room temperature and later resuspended in 100μl MilliQ water. The BAC DNA solutions were then put together in one eppendorf tube.

2.5.2. Isolation of BAC/PAC DNA for production of transgenic mice by pronuclear microinjection

After maxi-preparation of the BAC DNA, linearization and removal of the BAC vector started with a restriction assay. For human BAC RP11-626H4 (or human *PDGFRA* BAC-110kb), two restriction reactions each adding 5 μl of Not I restriction enzyme to 80 μl of BAC DNA, plus 20 μl restriction buffer (NE Biolabs), 20 μl BSA, and 75 μl ultrapure water to get a total 200μl reaction volume, incubated overnight at 37°C. These steps were done initially by Francoise Jamen. This procedure was then repeated by me to start the deletion of CS#1 (Fig 5.2, Chapter 5.2.3) from this human *PDGFRA* BAC-110kb by homologous recombination.

Later, the digested BAC DNA was run in a gel by PFGE (2.4.5.1.) with the program: 6V, 2.1-10 switch time, 6:30 hr in 0.5X TBE buffer; using NotI and undigested BAC DNA as control. Once the BAC DNA band was identified, its isolation continued identically to the YAC DNA purification procedure already described (2.4.5., Fig.2.1), with the minielectrophoresis (2.4.5.2., Fig. 2.1) run at 50V for 9 hours at 4°C. Later, the BAC DNA block was equilibrated with TENPA buffer (1M Tris-HCl pH 7.5, 0.5M EDTA pH 8.0, 5M NaCl) containing fresh 30μM spermine and 70μM spermidine, followed by agarose digestion and microdialysis (2.4.5.3. and 2.4.5.4., Fig. 2.1). However, the rough BAC DNA quantification was done by PFGE overnight, along with some previously isolated BAC DNA standard for comparison. The DNA was then centrifuged and diluted appropriately to a final ~1ng of linear BAC DNA in microinjection buffer. The pronuclear microinjections of human BAC RP11-626H4 DNA constructs were carried out by Francoise Jamen, Ulla Dennehy and Palma Iannarelli. They transferred 18-22 microinjected embryos per pseudopregnant mother.

2.5.3. BAC engineering by homologous recombination.

2.5.3.1. Construction of targeting vector

In order to construct the targeting vector to delete the CS#1 element (Fig. 5.2, Chapter 5.2.3) from the human *PDGFRA* BAC-110kb (Fig. 4.2) by homologous recombination. I designed two primers 80bp long with three main sequences, a restriction site, the homologous sequence next to the 5' and 3' sequence of the CS#1 element, and the homologous sequence to the pIGCN21 plasmid before its FRT sites (Lee et al., 2001). The

forward primer contained an AscI restriction site, 54bp 5'homologous sequence to the CS#1, and 5' homologous sequence to the pIGCN21 plasmid before its 5' FRT site; and the reverse primer starts with the 3' homologous sequence after the second FRT site of the pIGCN21 plasmid, 52 bp 3'homologous sequence of CS#1, and a PacI restriction site. The primers were ordered from MWG Biotech. Therefore the high fidelity PCR reaction (2.2.14) with these primers amplified the FRT-Kanamycin-FRT sequence from the pIGCN21 plasmid, as well as adding the 5' and 3' homologous sequence to CS#1 element and two restriction sites on either end to later ligate it to the pBluescript plasmid. The 1.5 kb PCR product was isolated from the agarose gel with the MinElute gel extraction kit (50) QIAGEN (Cat. No. 28604) or the GFX PCR DNA and gel band purification kit (Amersham Biosciences, product no. 27-9602-01). This pure PCR DNA was quantified with the Nanodrop spectrophotometer system. This pure high fidelity PCR product was used directly as targeting vector for the homologous recombination experiment.

Alternatively, this PCR product was also digested with AscI and PacI restriction enzymes to generate overhanging ends, purified again and quantified by Nanodrop. This was followed by a ligation reaction (2.2.8), consisting of 264.7 ng of this AscI/PacI digested targeting vector was mixed with 100 ng of linear pBluescript SK PacI/AscI digested plasmid (insert: vector molar ratio of 3:1), 1µl T4 ligase, 1 µl 10X T4 buffer and 0.5 µl 1mM ATP in a 10 µl total volume reaction, incubated at 16°C, overnight. After the ligation reaction, the DNA was purified again with the GFX PCR DNA and gel band purification kit (Amersham Biosciences). Transformation of XLBlue electrocompetent bacteria with this pure targeting vector-plasmid ligation was done as described in 2.1.2, and let to grow on a LB agar plate containing 100 µg/ml Ampicilin and 15µg/ml

Kanamycin. The selection of individual bacterial colonies started with mini-preparation of plasmid DNA (2.2.1) and then restriction assay using ClaI or AscI and PacI enzymes.

2.5.3.2. Electroporation of EL250 electrocompetent cells with the human *PDGFRA* BAC-110 kb.

The EL250 bacterial strain is a modified DH10B strain containing defective lambda prophage and arabinose *flpe* gene (Yu et al., 2000;Lee et al., 2001), in which homologous recombination is induced at 42°C and keeps stable at 32°C. Electroporation of the human BAC RP11-626H4 DNA (or human *PDGFRA* BAC-110kb) into the EL250 strain started with plating EL250 bacteria onto a Kanamycin plate and incubating it at 32°C overnight. Individual colonies were picked and grown in 5ml Kanamycin LB cultures at 32°C, overnight. 50 ml of LB media without antibiotic were inoculated with 1 ml of overnight culture and let grow around 2-3 hours at 32°C shaking at 200 rpm, until OD 600= 0.5-0.8. From now on the procedure must not take longer that 45 min to the electroporation step, and the bacteria are kept cool most of the time. After incubation, both cultures were cooled quickly in ice water for 5 mins by shaking by hand, and later spun at 2200 rpm (1000g) in the centrifuge at 4°C for 8 mins. In the cold room, the bacterial pellet was resuspended very gently in 1 ml of sterile water, transferred to 1.5ml eppendorf tubes, and washed three times with 1 ml sterile water, spun down for 20 seconds and resuspended very gently each time. After washing, the pellets were resuspended in 100 µl of sterile water. 50 µl of these bacteria were put in a fresh eppendorf tube and 1µl or 1:10 dilution of maxi-preparation of human *PDGFRA* BAC-110kb DNA (2.5.1) were added with cut-wider tips, mixed gently, and passed to an ice cold 0.1cm electroporation cuvette avoiding

bubbles. Water drops and any moisture on the cuvettes was cleaned thoroughly with a tissue. A pulse at 1.8kV, using a Bio-Rad gene pulser 25uF with pulse controller set to 200 Ohms (Ec1 program), was given to the bacteria and immediately 500µl of LB media added to help bacteria to recover at 32°C with shaking for 1.5-2 hrs. 1 and 10 µl of electroporated EL250 bacteria were spread on Kanamycin and Chloramphenicol LB agar plates. The bacteria that integrated the BAC grew in both Kanamycin and Chloramphenicol. Mini-preparation of BAC DNA was prepared from 8 individual colonies from the Chloramphenicol plate, but using cut tips for all the procedures. Restriction assay with AgeI and NotI restriction enzymes of each of the 8 colonies DNA was done to check the proper integration of the BAC DNA in the EL250 bacteria, comparing it with the original human *PDGFRA* BAC-110kb DNA PFGE restriction pattern.

2.5.3.3. Transformation of EL250 cells with targeting vector and induction of homologous recombination.

Transformation of the human *PDGFRA* BAC-110kb-EL250 bacterial strain with the targeting vector, started by plating these bacteria onto a Chloramphenicol plate and incubation at 32°C overnight. Again individual colonies were picked and grown in 5ml Chloramphenicol LB cultures at 32°C, overnight. 50 ml of LB media without antibiotic were inoculated with 1 ml of overnight culture and let grow around 2-3 hours at 32°C shaking at 200 rpm, until OD_{600}=0.5-0.8. This time in order to induce the homologous recombination, deleting the CS#1 element from the human *PDGFRA* BAC-110kb, two 50ml falcon tubes with 10 ml each from the culture were placed at different temperatures,

one at 42°C with shaking, which will be the homologous recombination induced culture; and the second one at 32°C, the non-induced culture, shaken for 15 min. After incubation, both cultures were cooled quickly by rotating them by hand in ice water for 5 mins, and later spun at 2200 rpm (1000g) in the centrifuge already at 4°C for 8 mins. In the cold room, the pellets were resuspended in 1 ml of sterile water very gently, transferred to 1.5ml eppendorf tubes, and washed three times with 1 ml sterile water, spinning down for 20 seconds and resuspending very gently each time. After washing, the pellets were resuspended in 100 µl of sterile water. To 50 µl of each induced and non-induced bacteria, 100-300 ng of pure linear targeting vector were added and mixed in cold and labelled eppendorf tubes set on ice. The cultures were then transferred to two separate ice cold 0.1cm electroporation cuvettes avoiding bubbles. Again the electroporation was done as previously specified (2.5.3.2), adding immediately after the pulse 500µl of LB media at 32°C and then shaking for 1.5-2 hrs. 1 and 10 µl of the electroporated induced and non-induced EL250 bacteria were spread on Kanamycin+Chloramphenicol LB agar plates. After mini-preparation of BAC DNA from 10 induced and non-induced colonies, the screening to find the BAC colony where the homologous recombination, and as a result the deletion of CS#1 sequence, occurred was done by PCR. Primers amplifying from outside the CS#1 element (5'-CTG GGG CAT GGG AGA GC-3', 5'-CTA TTC ATC GAG TGT TCA TTG G-3') and a 5' further primer (5'-ATT GGC GCG CCG CTA AAT CTG CTC CGC TTC AT) were used. Looking for a 1.5 kb PCR product showing the increase in sequence, from 223 bp of the CS#1 original sequence to 1.55 kb of the targeting vector in the BAC DNA after homologous recombination. The human *PDGFRA* BAC-110kb DNA and pIGCN21 plasmid DNA were used as controls for the PCR. Once

the CE human *PDGFRA* BAC-EL250 colony was selected, I continue with maxipreparation of BAC DNA (2.5.1.).

2.6. Production of YAC/BAC transgenic mice by pronuclear injection.

2.6.1. Preparation of pipettes and microinjection needles for pronuclear injection.

Three different types of pipettes were used during the transgenic procedure. They were called handling, holding and transfer pipettes. Most of these pipettes were made at least one day before the day of injection to avoid any inconvenience during the procedure. The handling and transfer pipettes were made from 1.0 mm diameter thin walled borosilicate glass capillaries (World Precision Instruments). These were pulled over the flame of a Bunsen burner and cut in the middle with a diamond point pencil. Those pipettes with around 150-180 μm tip diameter were selected to be handling pipettes. Those straight pipettes with 120-150 μm internal diameters were cut also straight at the tip and polished lightly at the edge in the flame to be used as transfer pipettes.

Holding pipettes were made also from 1.0 mm diameter thin walled borosilicate glass capillaries (World Precision Instruments). The tube was also pulled over a Bunsen burner flame. Keeping the tube vertical, it was cut in the middle with a diamond point pencil, those with a straight vertical tip and around 100-120 μm external diameter were selected. The next step was to melt the tip carefully using a microforge (Micro Instruments LTD) to

narrow down its opening to a 15 μm internal diameter. This small opening sucked the egg and holds it for microinjection. Independently, microinjection needles were made from 1.0 mm thin walled borosilicate glass capillaries with inner filament (World Precision Instruments). These were pulled by a micropuller (Flaming/Brown micropipette puller, Model P-97, Sutter Instrument. CO.), different pulling programs were tested until the precise thin and breakable needle was obtained, making it perfect for pronuclear injections (program: 275 Heat, 170 Pull, 50 Vel, 200 Time, 300 P). All the needles were carefully placed in a box protecting its sharp end and stored away from dust.

2.6.2. Procedure to super-ovulate female mice.

The procedure requires two sets of injections. In the first hormone injection, 4 week old mice females (CBAXC57BL/10 F1 hybrid) were injected with 5 iu of follicle stimulating hormone (FSH, Folligon). Forty six hours later, these female mice were injected with 5 iu of human chorulonic gonadotrophin (hCG, Chorulon). After a few minutes, they were mated 1:1 with fertile male studs (CBAXC57BL/10 F1 hybrid). All super-ovulated females had a plug next day.

2.6.3. Procedure to prepare pseudo-pregnant female mice for foster mothers.

Six to eight week old female mice presumably in oestrous were mated with vasectomized male studs (Albino strain). Preferably, 12 females were mated with 6 males (1:2 mating), at the same time the super-ovulated females were set to mate. Females that were plugged

the next day were then labelled as pseudo-pregnant females. These pseudo-pregnant females were ready for egg transfer that day.

2.6.4. Extraction of 0.5 day old fertile eggs and pronuclear injection.

Six superovulated plugged females (matings from C57BL/6JX CBA/J) were culled by CO_2 inhalation. Their ovaries and oviducts were extracted by lifting them by the fat on top of the ovary and cutting at the beginning of the uterus to keep the ampula intact (Nagy et al., 2003). The ovaries and oviducts were placed in M2 media (Sigma) in a 35 mm petri dish. The ampula was then separated under microscope (LEICA MZ6), getting rid of the ovary and the bit of uterus. Inside the swollen ampula were the eggs, which were then released only by cutting the oviduct with the forceps in M2 media. At this stage the eggs were surrounded by cumulus cells creating a sticky cluster.

The egg clusters were then transferred with a handling pipette to another 35 mm Petri dish with 3ml of M2 media and 90 µl of 10mg/ml hyaluronidase (in M2 media) and incubated 5 min at room temperature. The hyaluronidase enzyme separated the cumulus cells from the eggs, which were then carefully transferred and washed in fresh M2 media.

A selection of eggs was made under the microscope (LEICA MZ6), selecting the fertile eggs from the unfertile ones, choosing particularly those fertile eggs where both polar bodies and pronucleus were clearly visible. Eleven drops of M16 media (Sigma) were put asymmetrically on a 35 mm Petri dish and covered with mineral oil (M8410, Sigma). In

groups of 30 eggs per drop, these 0.5 day old fertile eggs were incubated for 1 hr at 37°C in 5% CO_2/ 95 % air and at 100 % humidity (HETO-HOLTEN Cell, House 170 Incubator), before microinjection. All the equipment and material for microinjection was prepared in this period of time.

2.6.5. Microinjection of YAC DNA into the pronucleus and transfer of microinjected eggs to a foster mother

Initially, the microinjection needle was filled with fresh YAC DNA purified and quantified the day before. Holding pipette and microinjection needle were set in the microinjector (PM1000 Cell Microinjector, MDI MicroData Instrument, Inc.; or IM 300 Microinjector, NARISHIGE Japan). On the microscope (LEICA DMIRB), 20 eggs were placed in a drop of M2 media. At least, 120 eggs per section were microinjected with the YAC DNA into one pronucleus. After one hour of incubation at 37°C in 5% CO_2/ 95 % air and at 100% humidity (HETO-HOLTEN Cell, House 170 Incubator), the eggs that survived the injection were separated in a 35 mm Petri dish containing M2 media. The eggs were then ready for transfer to a foster mother.

Pseudopregnant female mice were then anaesthetized with a combination of Hypnorm (fentanyl/fluanisone) and Hypnovel (midazolam) (Wolfensohn and Lloyd, 1998). After approximately 20 min and checking the complete effect of the anaesthetic, the surgery was initiated by cutting a 0.5 cm incision in the right low back skin and muscle, the fat on top the ovary was seen directly through the incision. Holding from the fat, the ovary and oviduct were taken out with forceps. A few minutes before starting the surgery, a transfer

pipette was filled consecutively with M2 media, a bubble of air, 30 microinjected eggs and a small extra bit of M2 media. Once the beginning of the oviduct was detected, the tip of the transfer pipette was inserted in and the 30 eggs were pushed inside it (Nagy et al., 2003). Once checking the microinjected eggs inside the ampula, all the reproductory tract was carefully placed back inside the body. The muscle was put together, cleaned and the skin was closed and joined with a small drop of biological glue.

Chapter 3.

Cranial neural crest and mesenchyme specific regulatory elements are located in the upstream sequence of the Human *PDGFRA* gene.

3.1 Introduction

3.1.1. *PDGFRA* expression in neural crest cells, mesenchyme progenitor cells, and their derivatives.

Neural crest cells (NCC) are migratory cells that originate at the dorsal side of the neural tube. These cells travel along the entire embryo and form connective tissue, bone, secretory cells, glia, peripheral neurons, some cardiac tissues and many other tissues. Neural crest cells appear very early during mammalian development and subsequently migrate and differentiate in response to extra-cellular factors including fibroblast growth factor, neurotrophins and many matrix macromolecules (Newgreen and Erickson, 1986;Perris et al., 1989).

The Pdgfra ligands, PdgfA and PdgfC, are highly expressed in neural crest target tissues in the mouse, like the epithelial lining of the branchial arches and branchial pouches. It has been suggested that these ligands might act as long-range migration cues or post-migratory signals for neural crest cells in the cranial region. The mutant mouse *patch* (*Pth*), which carries a deletion at the *Pdgfra* gene, has defects in the palatal closure and fusion, nasal and cardiac septation, and the development of several bone and cartilage structures. After

a complete study of *Pth* embryos, it was shown that the main reason for these defects was disrupted development of non-neuronal neural crest of cephalic and cardiac origin (Gruneberg and Truslove, 1960;Stephenson et al., 1991;Schatteman et al., 1992;Morrison-Graham et al., 1992). Therefore, it was concluded that in the mouse, Pdgfra is required cell autonomously for the development of a subset of non-neuronal neural crest cell derivatives in the cardiac and cranial regions (Morrison-Graham et al., 1992).

Later, four *PDGFRA* transgenic mouse lines were analysed and found to express their human *PDGFRA-LacZ* transgene specifically in mesoderm and neural crest-derived tissues (Reinertsen et al., 1997;Zhang et al., 1998). Reinertsen et al. (1997) identified *cis*-acting regulatory elements in the *PDGFRA* gene that lie within 6kb upstream of the transcription start site. They described expression of the *LacZ* reporter gene in several independent transgenic mouse strains, recapitulating predominant features of *PDGFRA* gene expression during mouse development, particularly in nervous tissue, skeleton, and several organ systems. Later, Zhang and collaborators tested three *PDGFRA* plasmid constructs containing 0.4kb, 0.9kb or 2.2 kb upstream sequence, plus the first exon of the human *PDGFRA* gene followed by the *LacZ* reporter. From the 2.2kb *PDGFRA-LacZ* transgene, 12.5-13 day old transgenic embryos showed expression in skeletal structures including the condensing skeletal primordia, perichondria and chondrocytes; the mesenchyme of hypodermis, dermal hair papillae, nostrils, lips, tongue, pharyngeal opening, trachea, lung, heart valves, pancreatic primordium, ureters, metanephros as well as cochlea, semicircular canal, and cerebellar anlage. Many of these cells could be neural crest cells or their derivatives, located already in the organs where they will differentiate.

In summary, it seemed that both the 0.9kb and 2.2 kb *PDGFRA-LacZ* transgenes could activate the expression of the gene in cranial neural crest cells and mesenchyme.

Other mutant mice that have *Pdgfra* conditionally deleted in neural crest cells showed several cardiac septation and remodelling defects similar to those observed in neural crest ablation studies. These results showed that the *Pdgfra* mutants are deficient for functional cardiac neural crest cells. However it is still not known whether such deficiencies are caused by malfunction or de-differentiation of these NC cells or but a reduced number of NC cells that reach target tissues (Tallquist and Soriano, 2003;Hoch and Soriano, 2003).

The roles of PDGF in neural crest are likely to be conserved among vertebrates, because in Xenopus and zebrafish embryos, *pdgfa* and *pdgfra* are similarly expressed in the pharyngeal region and neural crest, respectively.

3.1.2. YAC transgenic analysis of the specific regulatory elements of the human *PDGFRA* gene during development

The use of yeast artificial chromosomes (YACs), P1-bacterial artificial chromosomes (PACs) and bacterial artificial chromosomes (BACs) is one of the most powerful methods to analyze the expression pattern of vertebrate genes in different model organisms, from zebrafish to mouse. These artificial chromosomes allow cloning of long fragments of DNA, 50kb to 1 Mb long, which can contain complete gene(s) and inter-genomic sequences. These artificial chromosomes make possible the purification and even modification of the inserted DNA sequence. The DNA inserted can be purified and

microinjected into early embryos to produce transgenic animals, which can show partial or full expression of the endogenous gene.

YACs are linear yeast artificial chromosomes used as eukaryotic cloning vectors. The YACs are capable of maintaining stable DNA genomic fragments as long as 1Mb (Burke et al., 1987). They have a cloning site (SUP4) and two arms that contain essential yeast chromosome structures. The left YAC arm carries a telomere (TEL), a centromere (CEN4) and an autonomous replicating sequence (ARS1). The right YAC arm carries a telomere (TEL), and both arms can contain selectable markers, such as tryptophan (TRP) and uracil (URA) (Burke et al., 1987).

The first transgenic animal produced by pronuclear injection of a YAC was published in 1993 (Montoliu et al., 1993;Schedl et al., 1993a;Schedl et al., 1993b). The longest YAC that has shown authentic gene expression was a 680 kb Myf-5 YAC transgene; this study helped the authors to understand the regulatory control region of the Myf-5 gene and they described regulatory elements at a distance from the coding sequence of the gene (Zweigerdt et al., 1997). Because of their long inserts, often all the regulatory elements needed for the establishment and maintenance of the normal expression domain are located within the YAC, ensuring integration site-independent expression and often optimal levels of expression of the transgenes. Among the disadvantages are that the YAC DNA must be prepared intact and is liable to shearing or other breakage during preparation and injection, the purification protocol is long and technically difficult and their expression levels seem to be copy-number dependent (Huxley, 1998;Giraldo and Montoliu, 2001). The decision whether or not to use YAC transgenes depends on the

research project; the use of YACs seemed the only option to search for *cis*-regulatory elements of the human *PDGFRA* gene because some elements, for example the OLP element, could be located far upstream or downstream of the gene, making in total around 450 kb sequence to analyse which can only be cloned into a YAC. Later, the research project could be focused on the study of smaller regions of the human DNA sequence in bacterial vectors like PACs, BACs or even plasmid vectors, which are easier to work with. More information about PACs and BACs will be provided in the next chapter of this thesis.

3.2 Results

3.2.1. Characterization and purification of three constructs of the human *PDGFRA* YAC, Y29E11.

3.2.1.1. Characterization of Y29E11 by PCR and Southern blot

Yeast artificial chromosome 29E11 (Y29E11) has a human genomic DNA insert that contains the complete human *PDGFRA* gene and flanking sequences (Fig 3.1, A). It was isolated from the CEPH YAC library using primers specific for human *PDGFRA* exon 2 and exon 20 (Spritz et al., 1994). Dr. Tao Sun obtained this Y29E11 from the Genome Technology Center, Leiden, and transferred it into the yeast strain YLBW3 (Sun, 1999). This YLBW3 strain was genetically manipulated to provide an endogenous yeast chromosome-free zone "window" with the size range 295 kb to 590 kb (Hamer et al., 1995). This "window" facilitated identification of Y29E11 and its purification by pulsed field gel electrophoresis (PFGE), reducing in this way any contamination from endogenous yeast chromosomes (Fig 3.1, B).

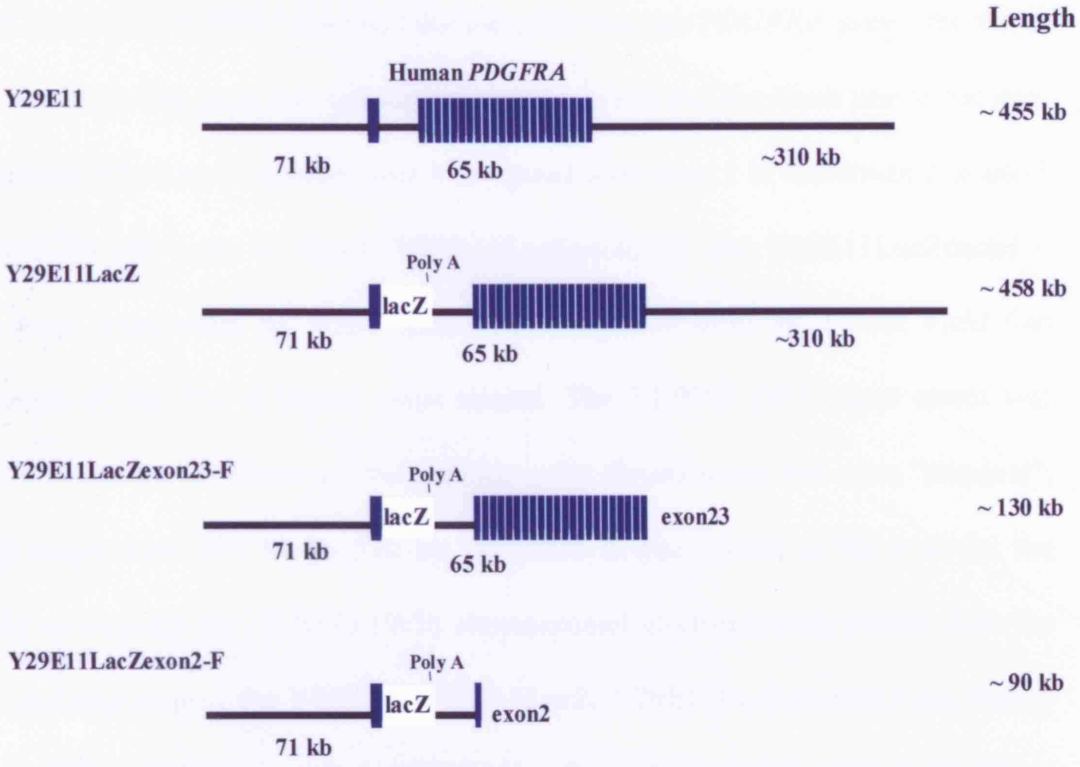
In order to check the integrity of the Y29E11, I performed PCR amplification of exon 1, promoter, and 3'-untranslated region (3'UTR) of the human *PDGFRA*, as well as the right and left YAC arms (see Chapter 2. 2.2.14, table 2.1 and 2.2). All the PCR results were positive and corresponded to the expected PCR products (data not shown). Therefore, the Y29E11 appeared to contain a complete copy of the human *PDGFRA* gene, within the limits of resolution of the PCR analysis.

In all vertebrates, *c-kit* gene is downstream of *pdgfra* gene in the chromosome cluster. It was important to determine whether the regulatory elements of the *KIT* gene could affect expression of *PDGFRA* in the Y29E11. Therefore, I also attempted to amplify the *KIT* promoter by PCR. There was no amplification from the human *KIT* promoter from Y29E11 DNA (Chapter 2, 2.2.14, table 2.1 and 2.2; data not shown). The lack of the *KIT* promoter reduces the risk of an indirect effect on *PDGFRA* gene regulation in Y29E11.

3.2.1.2. Culture and purification of the Y29E11 constructs.

Using the Y29E11, three more constructs were made by Vera van Limpt, Els Wessels and Dr. Gijs Afink (van Limpt, 1998;Wessels E., 2001). One construct Y29E11LacZ, contained an *nls-LacZ* reporter gene inserted at the end of *PDGFRA* exon1. In addition, they truncated this 450 kb YAC by homologous recombination in yeast. The truncation occurred at the end of exon 2 and exon 23 of the human *PDGFRA* gene, resulting in two YAC constructs named Y29E11LacZ-exon2-F and Y29E11LacZ-exon23-F, respectively

A Yeast artificial chromosomes (YAC):



B

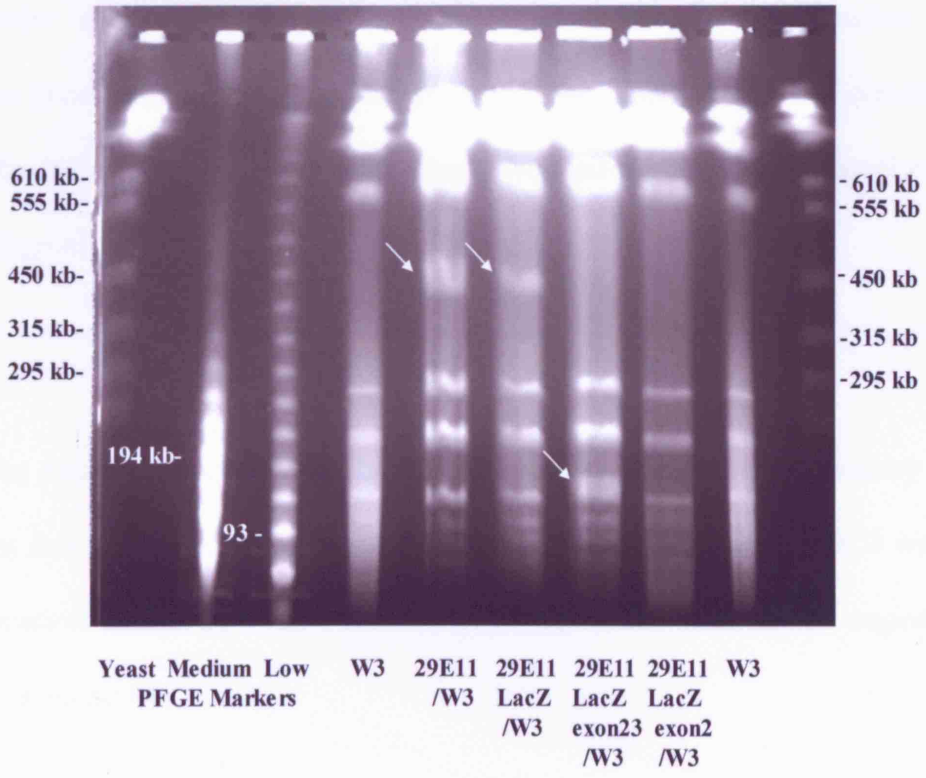


Fig 3.1. Y29E11 and its other three constructs to analyse the location of the Human *PDGFRA* gene regulatory elements. A. Schematic representation of the Y29E11 and fragmented constructs. Y29E11 contains the complete Human *PDGFRA* gene. The exons of the Human *PDGFRA* gene are represented as blue boxes and the black line is the non-coding sequence. The Lac Z reporter gene was ligated after exon 1 in constructs 2, 3 and 4 and contained its own poly A signal. Y29E11LacZexon23-F and Y29E11LacZexon2-F constructs were fragmented by yeast homologous recombination. **B.** Pulsed Field Gel Electrophoresis of the five different yeast strains. The YLBW3 (W3) yeast strain was genetically manipulated to show an endogenous yeast chromosome-free zone "window", with a size range from 295 kb to 590 kb, therefore it was chosen as the host for the Y29E11. By comparing the YLBW3 (W3) chromosomal electrophoresis pattern with the Y29E11 containing strains, the Y29E11, Y29E11LacZ, Y29E11LacZexon23-F are clearly shown in the PFGel (white arrows). YAC29E11LacZexon2-F is barely seen in this PFgel due to its smaller size in comparison with the other YACs, its identification and purification were done using a different PFGE program. The length of the YAC constructs was calculated by PFGE analysis using established low, medium and chromosomal range yeast markers (Sigma).

(Fig 3.1. A). The decision to work with the Y29E11 instead of the Y449C2 used by Sun et al. (2000) was due to technical reasons, the human *PDGFRA* insert in Y449C2 was in the wrong orientation in the YAC vector for truncations of the downstream region by homologous recombination.

The YAC purification procedures (described in Chapter 2, 2.4), begin with the culture of the yeast strain in selective culture media; the production of high density plugs containing the yeast DNA, separating the YAC from endogenous yeast chromosomes by PFGE, mini-electrophoresis to accumulate the YAC DNA into a small block of low melting point agarose; digestion of the agarose with the β-agarase; and finally microdialysis to transfer the vulnerable long YAC DNA into microinjection buffer solution. Each YAC strain was purified using a tailored PFGE program (details in Chapter 2, table 2.4)

An example of an analytical PFGE is shown in Fig 3.1. Southern blot analysis of the Y29E11 and Y29E11LacZexon2-F DNA was performed using a specific probe for the 3'-UTR human *PDGFRA* (Fig. 3.2). This confirmed that Y29E11 is ~450 kb long and that Y29E11LacZ-exon2-F is ~90 kb long and that Y29E11LacZ-exon23-F is ~130 kb long (Fig. 3.2. and not shown).

3.2.1.3. Obtaining the sequence of the Y29E11 insert ends using inverse PCR.

After the PFGE, Southern blot and PCR analysis, it was concluded that the Y29E11 is approximately 450 kb long and apparently contains the complete human *PDGFRA* gene. Considering that the Human *PDGFRA* gene is 65 kb long, the rest of the sequence in the Y29E11 insert had to consist of upstream and downstream sequences.

To determine the precise genomic endpoints of the Y29E11 insert and the lengths of 5' and 3' flanking sequences, I used inverse PCR to amplify the insert ends for DNA

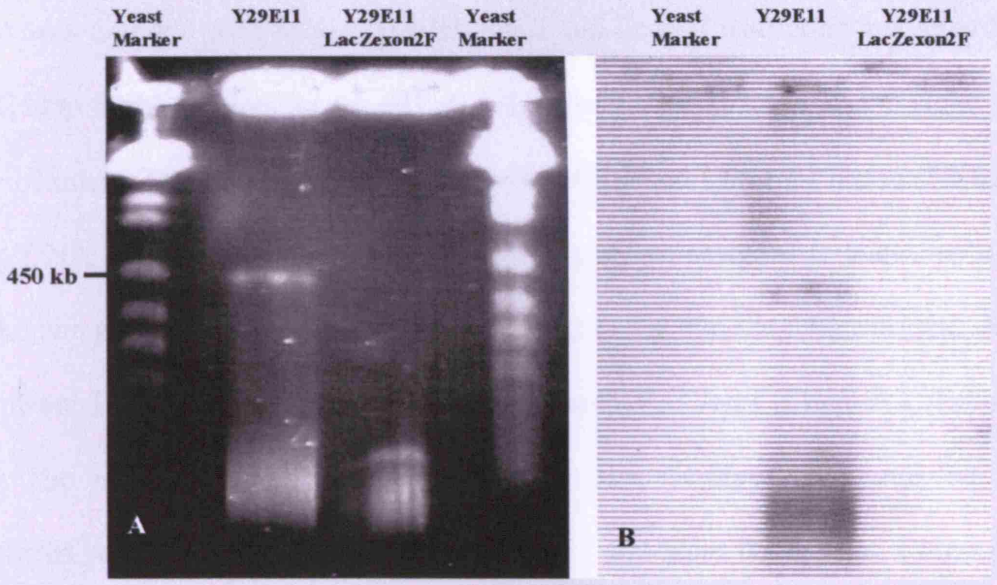


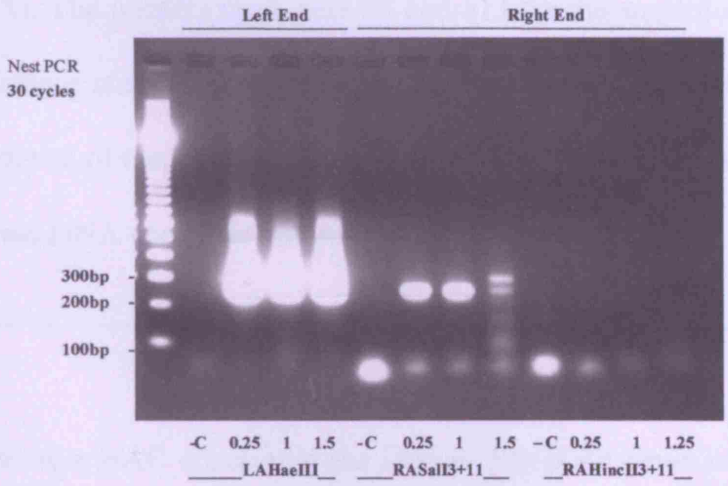
Fig 3.2. Pulsed field gel electrophoresis and Southern blot analysis of the Y29E11. A. PFGE of pure and intact Y29E11 DNA and Y29E11 Lac Z exon2-F DNA. B. Southern blot using the 3'UTR-human *PDGFRA* probe to check the integrity of the 450 kb Y29E11.

sequencing (Silverman G A, 1996). These sequences could be BLASTed in the whole available human genome sequence, to map the endpoints precisely. Purified Y29E11 DNA was digested with Mbo I, HaeIII, Rsa I and EcoRV restriction enzymes for the left YAC arm amplification, or HaeIII, Sal I, HincII, Sal I, Hinc II for right YAC arm amplification. The digestion fragments were self-ligated using T4 ligase (NEBioLabs) to form DNA circles. The inverse PCR amplification was designed to start amplifying from the known right or left YAC arm sequence, continuing into the unknown end sequence of the insert. I tried different sets of primers for each YAC arm (Chapter 2, Tables 2.1 and 2.2). The successful PCR reactions were with Hae III digestion of the left YAC arm amplified with primers 8 and 13, and with Sal I and Hinc II digestion of the right YAC arm, both amplified with primers 11 and 3. Nested PCR was performed to increase the concentration of these PCR products and ensure the specificity of the reaction (Fig 3.3. A). Each successful PCR product was purified from the electrophoresis gel using the Geneclean II kit, cloned into the p-GEM T vector for sequencing, using T7 sequencing primers (WIBR in-house sequencing facility). The sequences obtained are shown in Figure 3.3, B.

The amplified 223 bp insert sequence attached to the right YAC arm has not matched to any human genome sequence published to date. This was expected to match a sequence located around 310kb downstream of the Human *PDGFRA* gene. This could be because there is still a gap in the published sequence.

The amplified 143 bp sequence attached to the left YAC arm, matched with a sequence located 71082 bp upstream of the human *PDGFRA* gene. During this BLAST search I

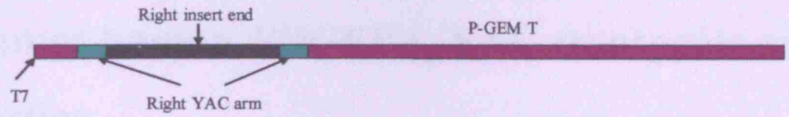
A



B

TCGCATGCTCCCGGCCGCCATGGTCCGCGGGATTGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGT
GGCTCCAAGTAGCGAAGCGAGCAAGACTGGGCGGCGGGTATACACTAGGTAACCCAGAACTCTCAGCCC
CAGGCAAGTGTAATATAAAGGTACATAGGATTGATTATGGAAACATTCTTAGAGGTGATGAGATTTGAGT
CTGGTTTTAAAATAGATGAGGTATAAGTGTATCTTTTAGAAGAATTCCGTAGTGATAAATTAAAGTCTTG
CGAATCACTAGTGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAG
TATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCT

Left end sequence of the Y29E11 Insert. This sequence belongs to the upstream sequence of *PDGFRA*.



NNANATTNNGCGGCCGACGTCGCTGCTCCCGGCCGCCATGNNCCGCGGGATTGCCCGATCTCAAGATTACGGA
CAGACAGACAGACAGACAGACAGACAGACAGACAGTACAGACAGACAGACAGACAGACAGACAGACAGACAGA
CAGACAGAGTTTGTCTTTGCTGTGTGATTCCCCTCAGGGTTCATTCTACCTGGAGAGAAAAGAAGTGAGAA
CTAGGAGACCGAAAGAATCCTAGCAAGTCGTCTCGCCCAGGAGACACCCACAGTCTCTCTCTTCTCTTTCT
ATGCTCTCTTTCCTCTCCCTTATGCGACTCTTAATCACTAGTGCGGCCGCCTGCAGGTCGACCATATGGGAGA
GCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATA
GCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC

Right end sequence of the 29E11 YAC Insert. This sequence should belong to the downstream sequence of *PDGFRA*, but no match has been found in the human genome database.

Fig 3.3. Inverse PCR experiment to rescue the end sequences of the Y29E11 insert. A. Agarose gel showing the inverse PCR products of self ligated-Y29E11 left and right end fragments, previously digested with Hae III for the left arm (LA), and Sal I and Hinc II to digest the right arm (RA). The primers used were #8 and #13 for the amplification of the left end self-ligated solution; and #3 and #11 for the amplification of the right end self-ligated solution. B. Sequence of the right and left end of the Y29E11 insert, and schematic representations of the final DNA contructs sent to be sequenced.

---

discovered the existence of a BAC containing the Human *PDGFRA* gene, named *homo sapiens* BAC clone RP11-601I15. This BAC is 130 kb long, contains up to exon 3 of the human *PDGFRA* gene and 97872 bp of upstream sequence (Chapter 4, Fig 4.1). Its complete sequence was published in January 2003. The comparison of the existing exon1 and exon2 sequences to this BAC sequence also helped to determine more accurately the length of intron 1 to 29341 bp. The analysis of the small left end sequence also confirmed the direction of the insert within Y29E11.

3.2.2. Production of human *PDGFRA* YAC transgenic mice by pronuclear injection.

I produced transgenic mice by microinjection of pure YAC DNA into pronuclei of 0.5 day mice fertile eggs (Fig 1.1B). Normally, six super-ovulated females produced at least 120 fertile eggs (The hormone injection procedure for the production of super-ovulated female mice and pseudo-pregnant female mice (CBAXC57BL/10 F1 hybrid) is described in

Chapter 2). I made the microinjection needles with a micropuller (Flaming/Brown micropipette puller, Model P-97, Sutter Instrument. CO). Handling and holding micropipettes by hand, polishing them in a microforge (Micro Instruments LTD). I always used fresh pure YAC DNA per microinjection session and injected at several different YAC DNA concentrations (PM1000 Cell Microinjector, MDI MicroData Instrument, Inc.; or IM 300 Microinjector, NARISHIGE Japan). Two hours of incubation after microinjection, I selected the surviving eggs and transferred them to the ampula at the oviduct of a pseudo-pregnant female (details in Chapter 2).

Later, 16 day old mouse pups were genotyped by PCR, amplifying either a 280 bp sequence of exon 1 of the human *PDGFRA* gene, or a 1095 bp sequence of the promoter region. A 884 bp sequence of the Lac Z reporter gene was also used to genotype three of the human *PDGFRA* YAC constructs (Fig 3.4). In addition, a probe for the 3'UTR-human *PDGFRA* gene was used in Southern blot analysis to confirm the presence of the Y29E11 in the genomic DNA.

After many months of YAC DNA purifications, pronuclear microinjection sessions and genotyping of embryos and pups, one transgenic founder was detected from Y29E11 (Fig 3.4. A), none from Y29E11LacZ, two transgenic founders from the Y29E11LacZexon2-F (Fig 3.4. B) and one founder from the Y29E11LacZexon23-F. The transgenes in each transgenic founder were characterized further to determine their integrity by amplifying different parts of their sequence as described in the next section.

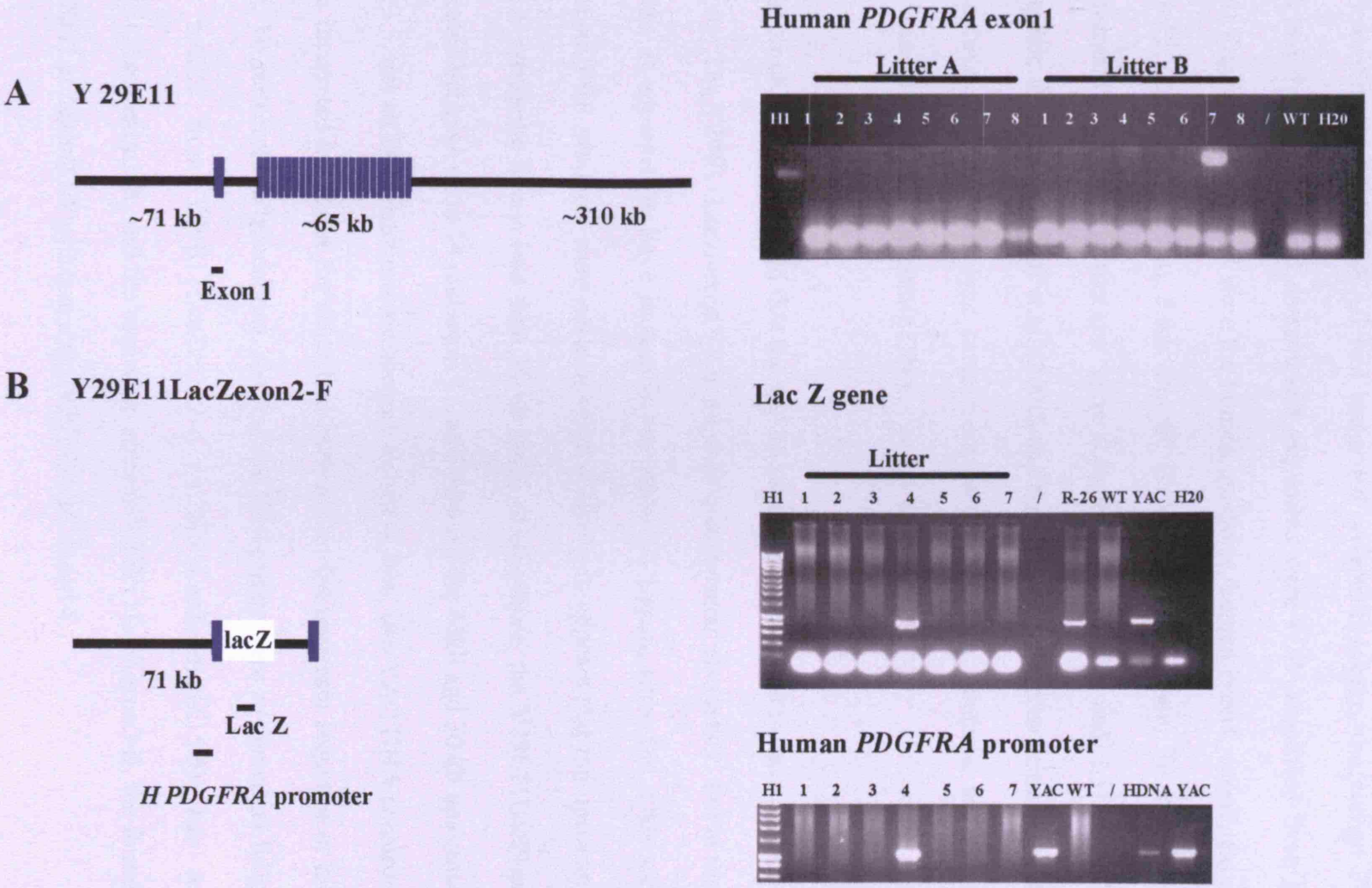


Fig 3.4. Genotyping of the Human *PDGFRA* YAC transgenes in mouse genomic DNA by PCR. A. Detection of a founder of the Y29E11 by PCR of the Human *PDGFRA* exon1 sequence. B. Detection of an E15.5 transgenic embryo from the Y29EllLacZexon2-F by PCR of the LacZ gene and the Human *PDGFRA* promoter sequence.

3.2.3. Characterization of the YAC transgene integrity by PCR

To determine whether the YAC had integrated without breakage, the human *PDGFRA* gene and the upstream and downstream sequences were PCR- amplified from genomic DNA. Fragments amplified were the human promoter, human exon1, exon2, exon 13-14, exon 17, exon 23, 3'UTR; 30kb and 60 kb upstream sequence; 70 kb and 105 kb downstream and Lac Z reporter gene (Chapter 2, table 2.2 and 2.3). The Y449C2 transgenic line made by Sun et at. (2000) and the Rosa26-lac Z transgenic line were used as positive controls; wild type mice were used as negative controls, as well as PCR reactions without added genomic DNA.

These amplifications showed that the 450 kb long Y29E11 seemed to have integrated in its entirety. The Y29E11LacZ-exon2-F in the first founder seem also intact, but in the second founder it appeared to have broken in the upstream region, since the 30kb and 60 kb upstream PCR products were not amplified. It therefore appears that the upstream region in this transgenic line is less than 30 kb long. Furthermore, the Y29E11LacZ-exon23-F truncated between exon 14 and exon 17, and between the 60kb and 30 kb upstream region (Fig 3.5, not all PCR reactions are shown). In conclusion, two YAC DNA constructs seem to be integrated intact but the other two truncated in the upstream sequence or inside the gene. To reflect these truncations, the founder transgenics were re-named as follows: the first mouse from Y29E11LacZexon2-F (Y29E11LacZexon2-E15.5) was re-named Y29E11LacZexon2-A, and the second founder to Y29E11LacZexon2-B, the founder from Y29E11LacZexon23-F was re-named Y29E11LacZexon14.

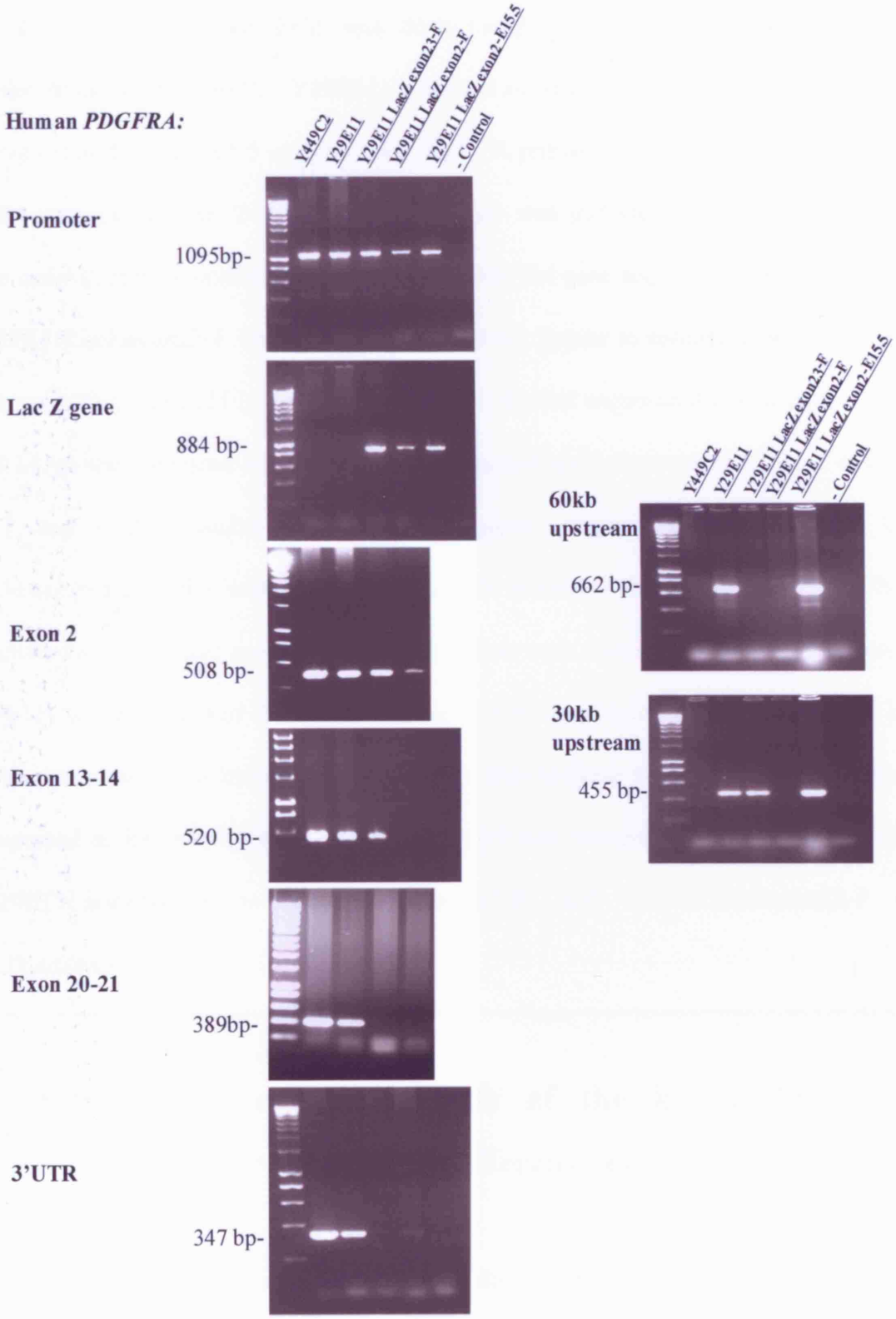

Human *PDGFRA:*
Y449C2
Y29E11
Y29E11 LacZ exon23-F
Y29E11 LacZ exon2-F
Y29E11 LacZ exon2-E15.5
- Control
Promoter
1095bp-
Lac Z gene
884 bp-
Exon 2
508 bp-
Exon 13-14
520 bp-
Exon 20-21
389bp-
3'UTR
347 bp-
60kb upstream
662 bp-
30kb upstream
455 bp-

Fig 3.5. Characterization of *PDGFRA* YAC transgenic founders. PCR amplifications of the human *PDGFRA* promoter, lacZ reporter gene, exon 2, exon 13-14, exon20-21, and the 3'UTR are shown. Each PCR was done using genomic DNA from the mouse transgenic founder of Y449C2, Y29E11, Y29E11LacZexon23-F, Y29E11LacZexon2-F, and Y29E11LacZexon2-E15.5 embryo. All the PCR primers were specific to the Human *PDGFRA* sequences. The Y449C2 transgenic line was included in this analysis as a positive control, since it contains all the human *PDGFRA* gene sequences, but lacks LacZ. The Y29E11LacZexon23-F transgenic founder did not appear to contain exon 17-18 (not shown), exon20-21, exon23 (not shown) or 3'UTR; the last sequence that it amplified was exon13-14, so the transgene was truncated after exon14 and before exon17. The Y449C2, Y29E11, and Y29E11LacZexon2-F transgenes seem complete. Only exon2 from the Y29E11LacZexon2-E15.5 was not shown. A PCR reaction without any genomic DNA was included as a negative control. Other truncations were seen at the upstream sequence of Y29E11LacZexon2-F and Y29E11LacZexon23-F by amplifying small sequences 30 kb and 60kb upstream of the human *PDGFRA* gene. Due to these truncations the transgenes were renamed as follows, Y29E11LacZexon2-E15.5 was changed to Y29E11LacZexon2-A, Y29E11LacZexon2-F to Y29E11LacZexon2-B, and Y29E11LacZexon23-F to Y29E11LacZexon14.

3.2.4. Expression pattern analysis of the human *PDGFRA* exon2-F YAC transgenic mice at different developmental stages

The expression pattern analysis started with the study of the Y29E11 transgenic. Expression of the YAC was checked in transverse sections of P0 transgenic pups by in situ

hybridization using a 3'UTR-human *PDGFRA* probe, which detects only human transcripts, not endogenous (mouse) transcripts. This probe was used previously when analysing Y449C2 transgenics (Sun et al., 2000). In addition, I used Y449C2 transgenic mice as a positive control. However, the Y29E11 line did not show any expression in any tissue (data not shown). The reason(s) why this YAC did not express is unknown. The transgenic line seems to contain all the 450 kb Y29E11 fragments amplified by PCR (Fig 3.5). Comparing the PCR products from the Y449C2 line with Y29E11 mice, it seems that there were less transgene copies in 29E11 than in 449C2 line. However, this does not explain the lack of expression.

Y29E11LacZexon2-F was integrated in two independent mice. The first was culled at embryonic day 15.5 (E15.5) and so a line was not derived from this mouse. The second mouse was established as a founder of the transgenic line to check at different embryonic stages. Both founders contained a YAC transgene that was truncated, the name given to these transgenes per founder are Y29E11LacZexon2-A and Y29E11LacZexon2-B respectively.

The Y29E11LacZexon2-A transgene was expressed in motor neurons and hair follicles as clearly shown by in situ hybridization with a Lac Z probe in transverse sections of the E15.5 embryo (Fig. 3.6. A, B and C). Expression was also evident in the mesenchyme of the nostrils, tongue, eye lids, jaw articulation and vibrissa hair bulbs by β-gal staining of coronal sections of the E15.5 embryos head (Fig. 3.6, D and E). In the brain, there was expression in some cells in the hippocampus, dentate gyrus, ventral periaqueductal gray matter, choroid plexus, cerebellum and medulla (Fig. 3.6, F, G and H).

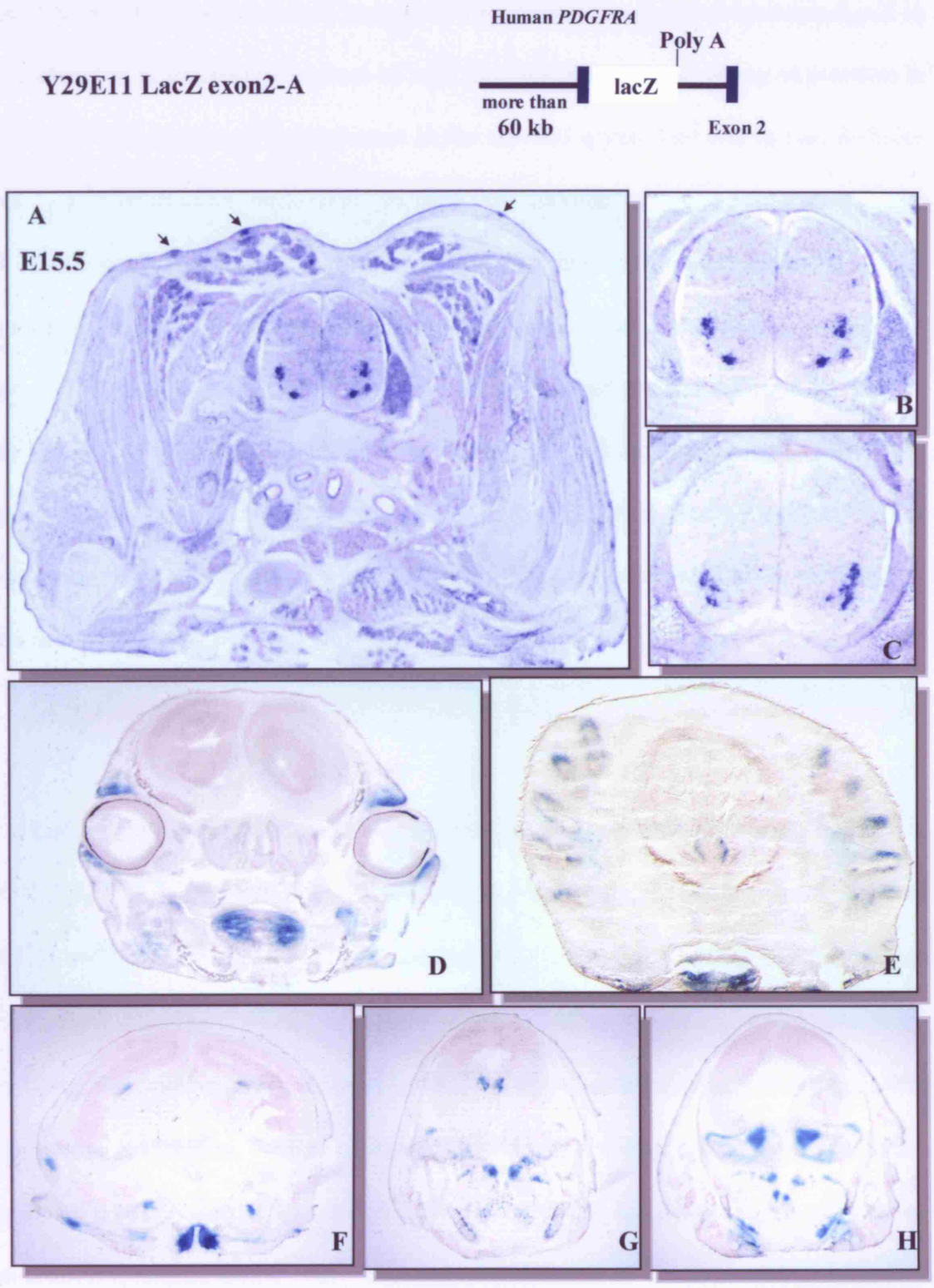

Human *PDGFRA*
Poly A
Y29E11 LacZ exon2-A
lacZ
more than
60 kb
Exon 2
A
E15.5
B
C
D
E
F
G
H

Fig 3.6. The Y29E11LacZexon2-A transgene expression in an E15.5 embryo. LacZ in situ hybridization in transverse sections of an E15.5 embryo, showed strong expression in what seems to be a cluster of motoneurons in the cervical spinal cord and in hair follicles (arrows, A). Consecutive transverse sections of cervical spinal cord showing the Y29E11LacZexon2-A transgene expression in different subsets of motoneurons (B, C). β-galactosidase staining of coronal sections of the head showed the expression of the transgene in the mesenchyme of the eye lids, the tongue and the jaw articulation (D), as well as in the hair follicles of the whiskers, the nostrils and again the tongue in a more frontal section (E). In the brain, there was expression in some clusters of cells in the hippocampus, dentate gyrus (F), ventral periaqueductal gray, choroid plexus, cerebellum, and medulla (G, H).

Y29E11LacZexon2-B transgenic line was analysed at pre-implantation stages, E10, E11, E11.5, E12.5, E13.5, E15.5 and P1. At E10, there was strong expression in the somites (Fig 3.8, B and C). In wholemount β-gal staining of E11, E11.5, E12.5 and E13.5 embryos (Fig. 3.7), the Y29E11LacZexon2-B is expressed in neural crest cells and sclerotome at all stages. Using the β-gal staining, it is not possible to distinguish between neural crest cells and sclerotome progenitors, but due to their location and number of cells it seems that both cell types are labelled. At E11 and E11.5 there is very strong LacZ expression in first and second branchial arches, neural crest cells and sclerotome (Fig. 3.7, E11 and E11.5). The neural crest cells migrating to several organs are very clear at these stages in cryostat sections (Fig. 3.8, A). At E12.5, there are now labelled cells in the extremities of the embryo at the level of bone articulations (Fig. 3.7, E12.5). At this stage, there are also

strongly labelled bone cells in the vertebrate (Fig 3.8. G) and in the mesenchyme of the kidneys, intestine and skin (Fig. 3.8, D, E and F). At E13.5, there was now clear expression in the mesenchyme of the eye lids, tongue, kidneys, intestine, and skin; there was also strong expression in the hair follicles, as clearly shown in the whiskers, and also in chondrocytes of the vertebrae and ribs (Fig 3.7, E13.5, and data not shown). This entire expression pattern was similar to the one shown by the E15.5 embryo with the Y29E11LacZexon2-A transgene (Fig. 3.6).

Special attention was paid to the expression pattern in the spinal cord. At E11 there was only expression in some cells clustered at the site of motoneurons. At E13.5, the expression was just restricted to four or five cells per section in the ventral spinal cord, and later in development there was no expression anywhere in the spinal cord (Fig. 3.9). In the brain, there was expression only at E13.5 in a subset of cells (presumably neurons or precursors) in the hypothalamus and a few cells in the corpus striatum mediale (Fig 3.8, H and I). These results, the endogenous mouse *Pdgfra* gene expression and the Y449C2 expression pattern (Sun et al., 2000) are summarised in table 3.1. Most of the expression pattern described was in cells already described to express the endogenous *Pdgfra* gene, except for motoneurons (table 3.3).

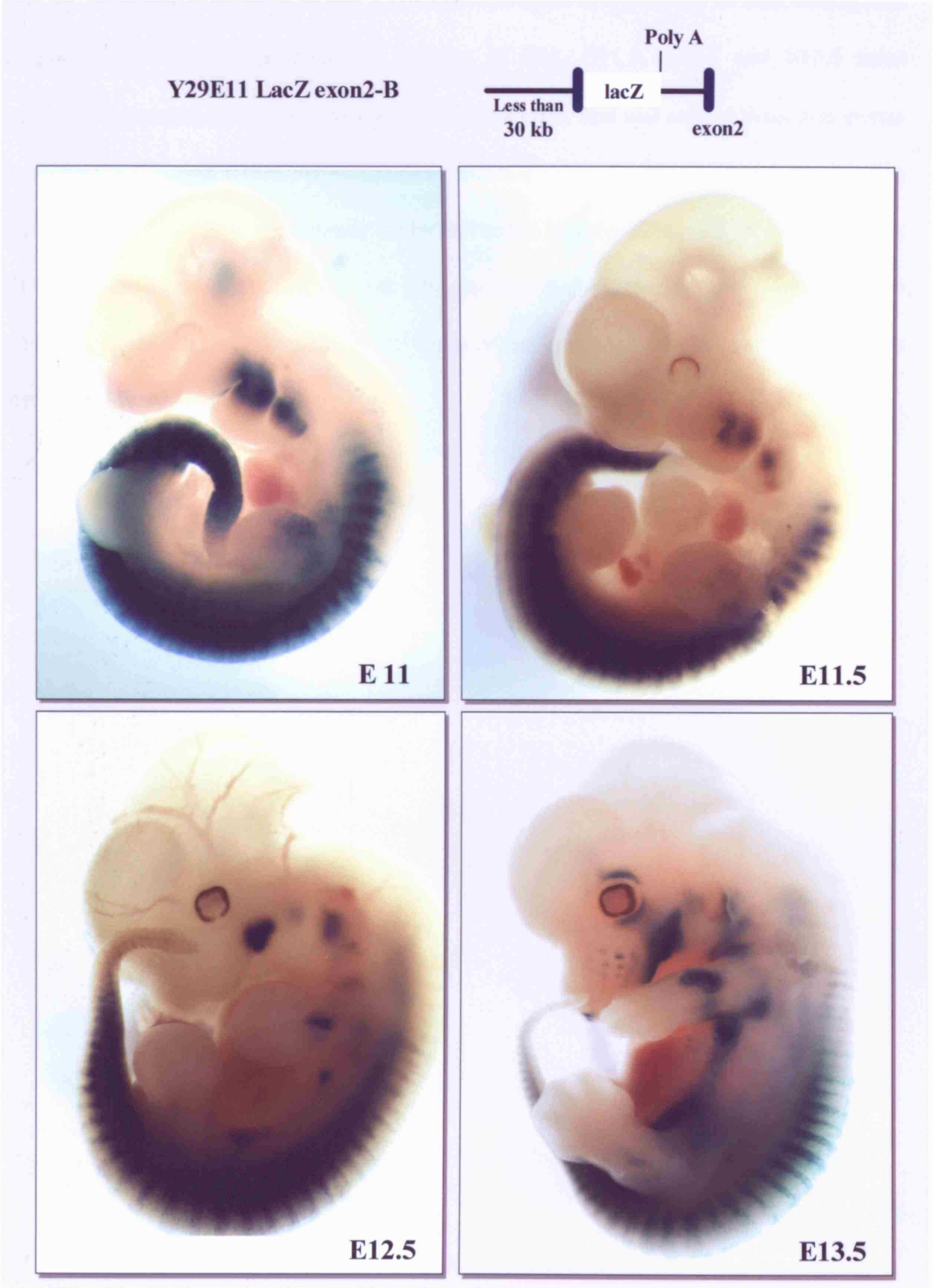

Y29E11 LacZ exon2-B
Poly A
Less than
30 kb
lacZ
exon2
E 11
E11.5
E12.5
E13.5

Fig 3.7. Expression pattern of the Y29E11LacZexon2-B transgene at mid embryonic stages. Wholemount β-galactosidase staining of E11, E11.5, E12.5 and E13.5 mice embryos. Expression in the sclerotome, neural crest cells, first and second branchial arches is evident at E11 and E11.5. At E12.5, there are now strong positive cells clustered in the jaw articulation, cells in the anterior limbs that could be osteoblasts and or chondrocytes. At E13.5, the mesenchyme of the eye lids, jaw articulation, tongue, hair follicles of the whiskers, nostrils, cluster of cells in the bones of the limbs and skin cells in the abdomen are all positively labelled.

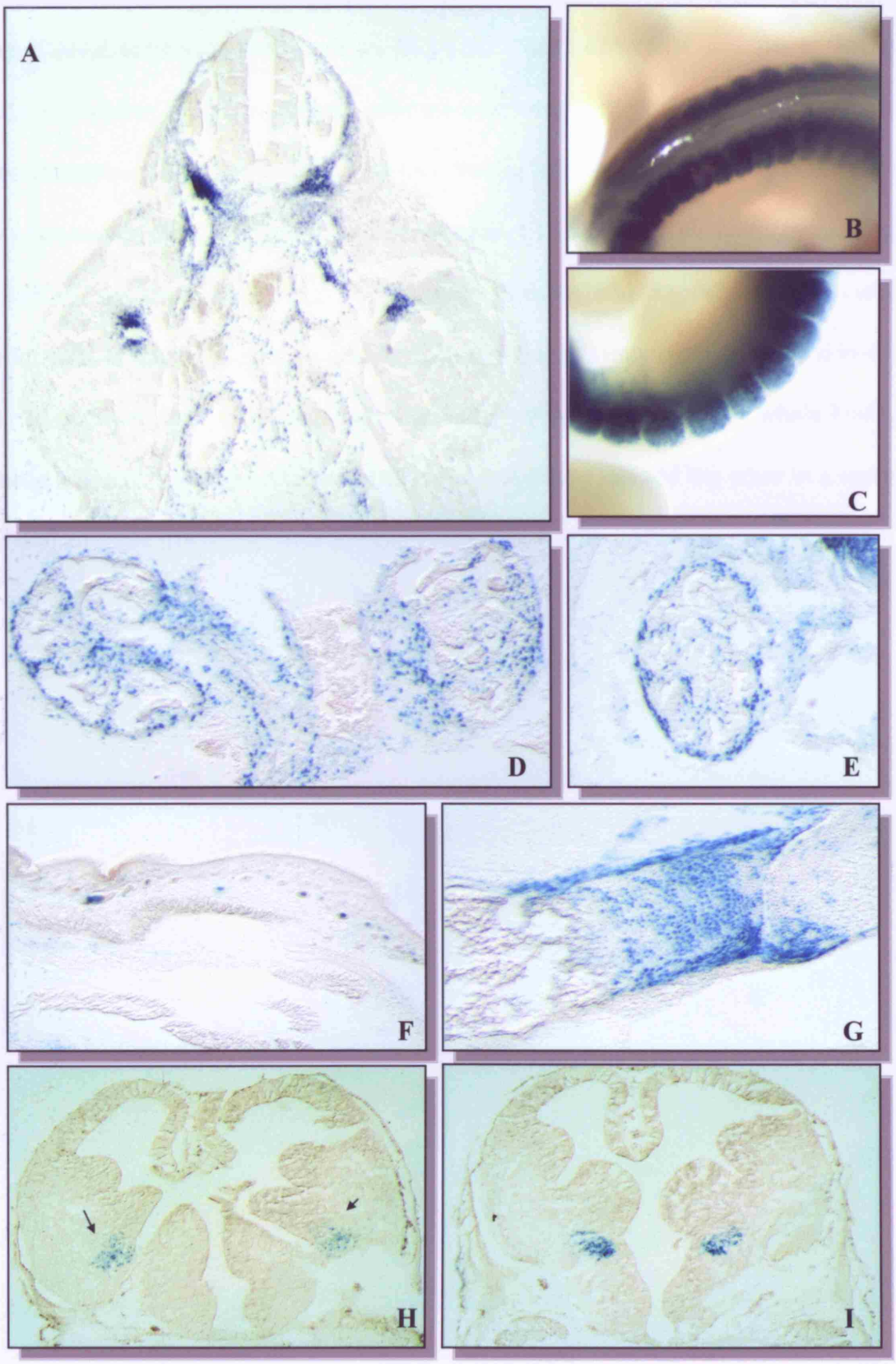

A
B
C
D
E
F
G
H
I

Fig 3.8. Study of the Y29E11LacZexon2-B transgene expression by β-galactosidase labelling. Caudal transverse section of an E12-12.5 mice embryo. At this stage, neural crest cells, sclerotome and mesenchymal cells are seen migrating ventrally to other organs and bone formation of the lower limb bud (A). Strong expression in somites of an E10 embryo wholemount (B, C). Transverse sections of an E13.5 embryo showing expression in the kidney mesenchyme (D,E), dermal hair follicles and hypodermis (F), and sclerotome cells in the vertebrae (G) possibly chondrocytes. Coronal sections of the brain of an E13.5 embryo, there was LacZ expression only in two sections of the whole brain, one showed a few cells in the corpus striatum mediale (arrows, H) and the other in a more compact and strongly labelled cluster of cells in the hypothalamus (I).

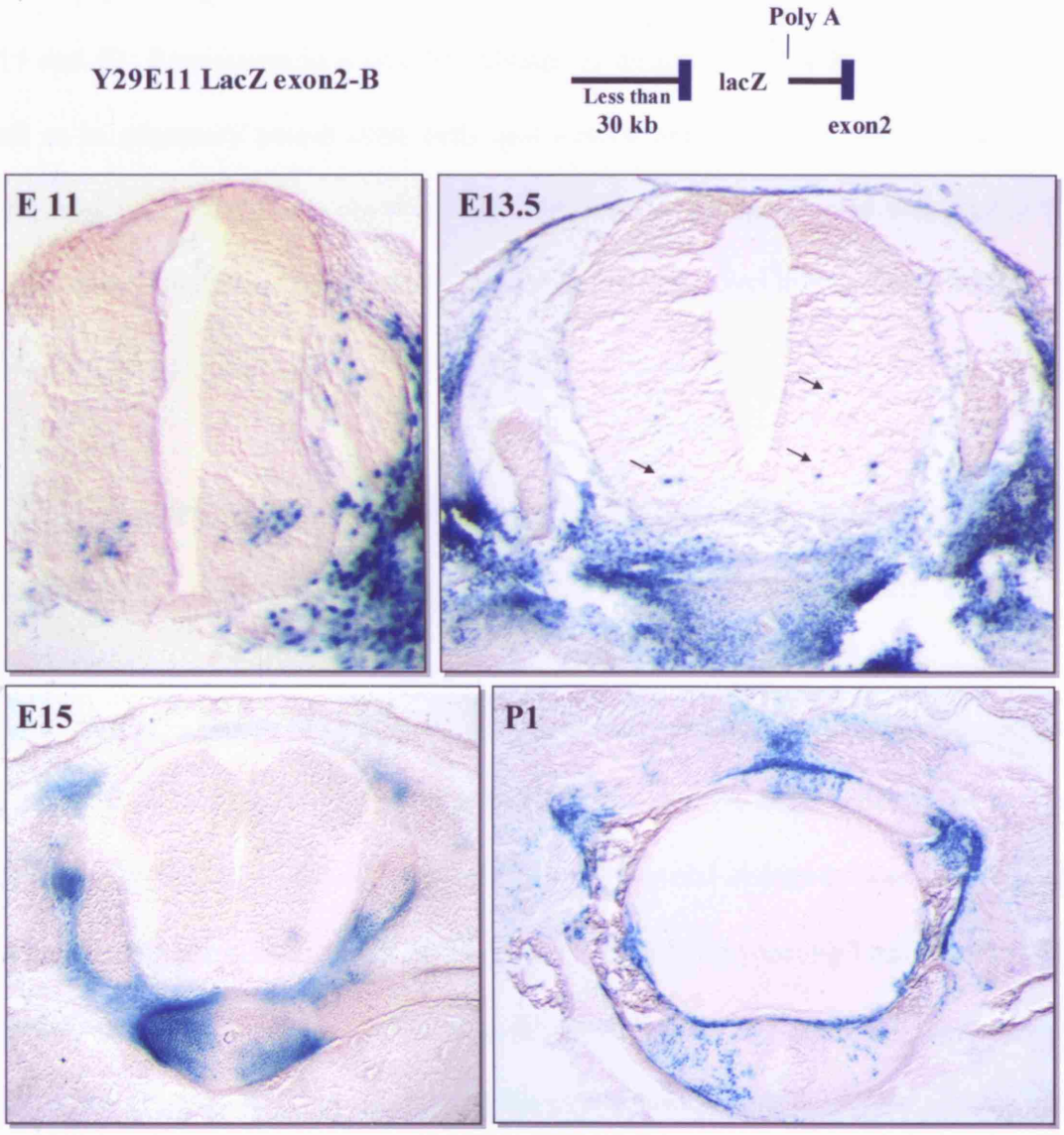

Y29E11 LacZ exon2-B
Poly A
Less than
30 kb
lacZ
exon2
E 11
E13.5
E15
P1

Fig 3.9. Analysis of the expression pattern of the Y29E11LacZexon2-B transgene in the embryonic spinal cord. Transverse sections of cervical spinal cord at E11, E13.5, E15 and P1. Expression in a possible cluster of motoneurons at E11 was very strong, as well as in migratory neural crest cells and sclerotome. At E13.5 only a few cells in the ventral spinal cord are seen (arrows). No expression in the spinal cord was seen at E15 and P1, but there was still expression in sclerotome cells involved in vertebrae formation.

3.2.5. Expression pattern analysis of the human *PDGFRA* exon14-F YAC transgene at different developmental stages

The Y29E11LacZexon14 transgenic line was analysed at pre-implantation stages, E11, E12.5, E13, E14.5 and P2. It is expressed in neural crest cells, sclerotome progenitors, trigeminal ganglion cells, and first and second branchial arches as was shown clearly by wholemount β-gal staining (Fig. 3.10). At E11 there is very strong LacZ expression in the first branchial arch, only a few cells in the second branchial arch, neural crest cells and sclerotome (Fig. 3.10, E11). At E12.5, the expression pattern was very similar to that at E11, except for additional strong expression in the dorsal root ganglia (Fig. 3.10, E12.5 and Fig. 3.12, E12.5). In transverse sections of these E11 and E12.5 embryos, some neural crest and sclerotome cells seem to be migrating ventrally to other organs in the embryo (Fig. 3.11, A, B and C). At E13-13.5, there was now clear expression in the mesenchyme of the jaw articulation (Fig 3.10, E13 and Fig 3.11, B). At E14.5, there was additional expression in the mesenchyme of the eye lids and tongue (Fig 3.10. E14.5). At P2, there was still strong expression around the vertebrae, in bone cells (possibly chondrocytes),

DRGs, neurons at the dorsal part of the spinal cord (Fig 3.11, D; Fig 3.12, P2) and the mesenchyme of the intestine (Fig 3.11, E).

The spinal cord at E11 showed expression in a few cells at the level of the motoneurons and many others distributed along the dorsal to ventral axis of the spinal cord (Fig. 3.12, E11), at this point there was also expression in the sclerotome surrounding the neural tube. At E12.5, there was strong expression in the DRGs and many cells in the ventral part of the spinal cord in a V shape distribution, outside the ventricular zone (Fig 3.12. E12.5). At E13, strong expression was detected in many cells in the dorsal spinal cord and only four or five cells per section in a small cluster in the ventral region (Fig. 3.12, E13). At P2, the expression in the dorsal spinal cord is still in many cells but no cells are labelled in the ventral part. The DRGs are strongly expressed in the Y29E11LacZexon14 from E12.5 until P2 (Fig 3.11, A, B, C and D, Fig 3.12).

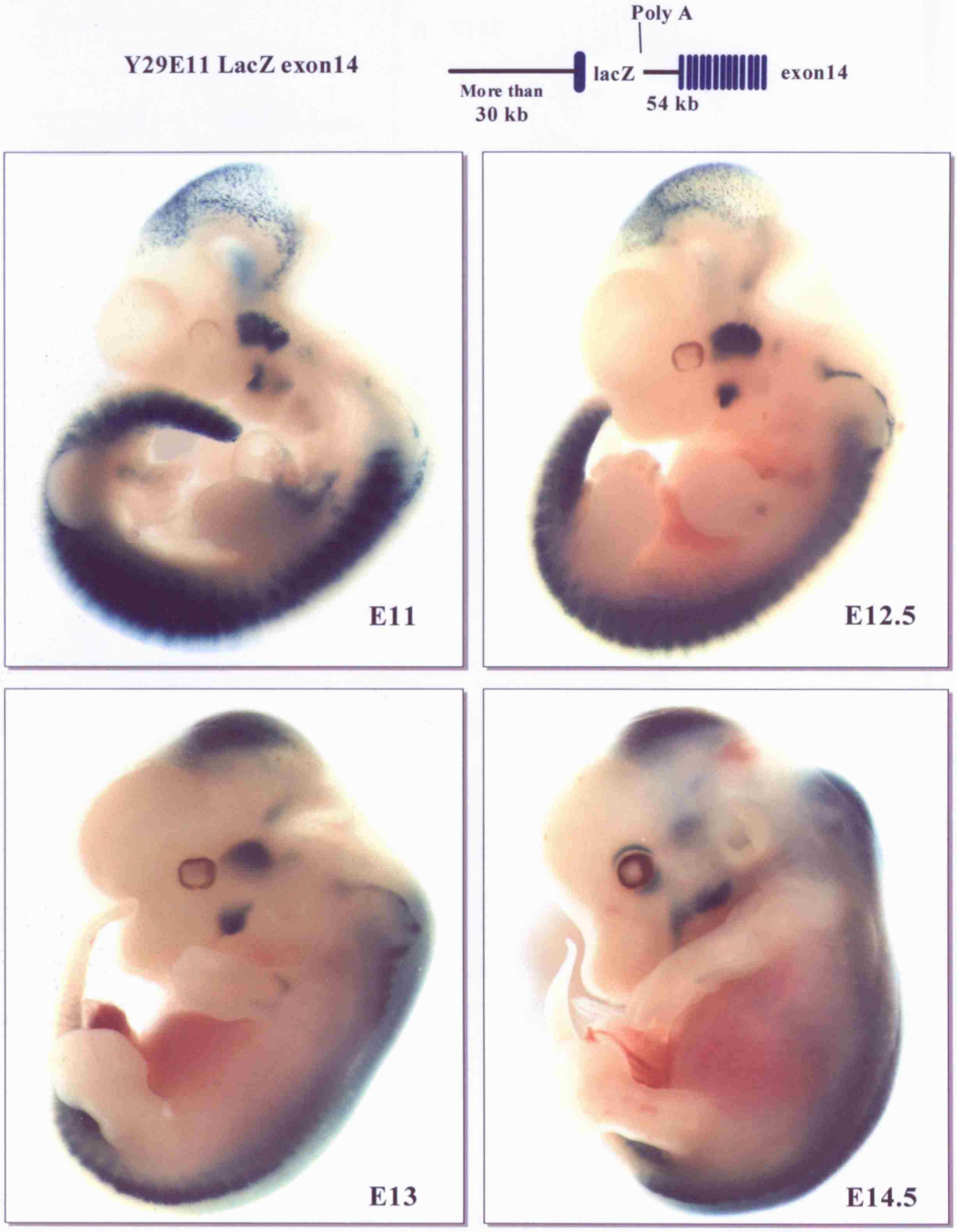
In the brain, there was Y29E11LacZexon14 expression in some cells at the periphery of the midbrain throughout development (Fig 3.10. and Fig 3.11, G), and even at P8 the expression in the midbrain was still strong (Fig. 3.11, J). There was also expression in some cells arranged in a line in the diencephalic region (arrow), and spread in the midbrain at E13 (Fig. 3.11, H and I), no more labelled brain cells were found in any other embryonic stages. Furthermore, the expression at the trigeminal ganglion cells is strong and evident from E11 to E14.5 (Fig 3.10, Fig 3.11, F).

In general, the expression of Y29E11LacZexon14 is less wide spread than Y29E11LacZexon2-B, most of the labelled cells seem to be in and around the spinal cord

and vertebral column (Fig. 3.11, B and C), only the mesenchyme of the intestine shows expression at P2 as clearly shown in table 3.2, and no other important organs show any expression during development.

The β-gal wholemount expression pattern at E11, E12.5 and E13.5 embryonic stages of this transgenic line as well as Y29E11LacZexon2B, were later compared with the previously published 2.2 kb LacZ transgenic line, which embryos were kindly donated by Zhang and β-gal stained following my method for accurate comparison (Zhang et al., 1998) (3.3.4. and Fig 3.14). Additionally, these human *PDGFRA* YACs, 2.2kb human *PDGFRA-lacZ* (Zhang et al., 1998), 6kb human *PDGFRA-lacZ* (Reinertsen et al., 1997), Y449C2 (Sun et al., 2000) expression patterns and endogenous mouse *Pdgfra* are summarised in table 3.3.

Fig 3.10. Y29E11LacZexon14 transgene expression in four embryonic stages. Wholemount β-gal staining of E11, E12.5, E13.5 and E14.5 embryos. At E11, strong expression was seen in the trigeminal ganglion, DRGs, widely distributed cells in the midbrain, first branchial arch, neural crest cells and sclerotome progenitors. At E12.5 and E13, kept the same expression in the same cells but there were now clusters of cells in the jaw articulation, in bone formation of the forearms, and mesenchymal cells migrating ventrally in the hypodermis. At E14.5, there was also expression in the mesenchyme of the eye lids, jaw and tongue.


Y29E11 LacZ exon14
Poly A
lacZ
exon14
More than
30 kb
54 kb
E11
E12.5
E13
E14.5

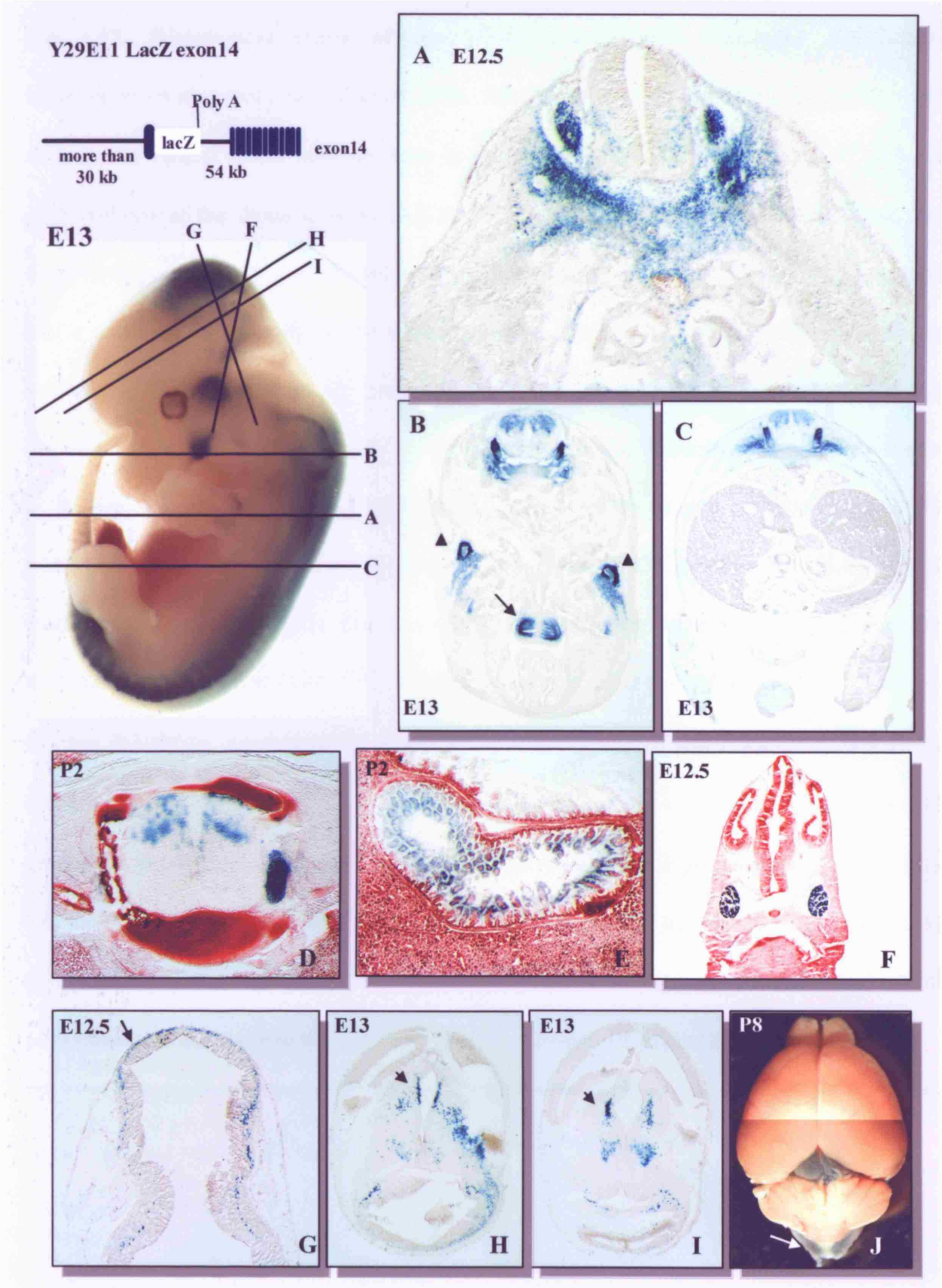

Y29E11 LacZ exon14
Poly A
more than 30 kb
lacZ
54 kb
exon14
E13
G
F
H
I
B
A
C
A E12.5
B
E13
C
E13
P2
D
P2
E
E12.5
F
E12.5
G
E13
H
E13
I
P8
J

Fig 3.11. Histological study of the Y29E11LacZexon14 transgene expression. Expression in migratory neural crest cells, sclerotome, dorsal root ganglia (DRGs) and cells in the ventral spinal cord are seen in a transversal section β-gal stained of an E12-12.5 embryo at the thoracic level (A). At E13 in a transverse section of the head, the expression pattern in the spinal cord changes dramatically, mainly in the dorsal spinal cord and a small cluster of cells in the ventral domain, there is also strong expression in the sclerotome of the vertebrae, the mesenchyme at the jaw articulation (arrowhead, B) and the mesenchyme of the tongue (arrow, B). A section at the thoracic-abdominal level also at E13 shows no expression in the lungs, liver, intestine or any other organs(C). P2 transverse sections showed continuing expression in the DRGs, in some cells in the dorsal spinal cord mainly located in the grey matter (D), and now there was expression in the mesenchyme of the intestine (E). A coronal section of the head of an E12.5 embryo showed the strong staining at the trigeminal ganglia (F). D, E and F sections were β-gal and neutral fast red stained. E12.5 brain coronal sections showed positive β-gal cells all around the most external layer of the midbrain (arrow), anterior pons and cerebellum (G). Transverse sections of the E13 brain showed clusters of cells labelled in the diencephalic region (arrows) and midbrain (H and I). Clear expression in the midbrain and dorsal spinal cord (arrow) is still seen in the wholemount β-gal stained P8 brain (J).

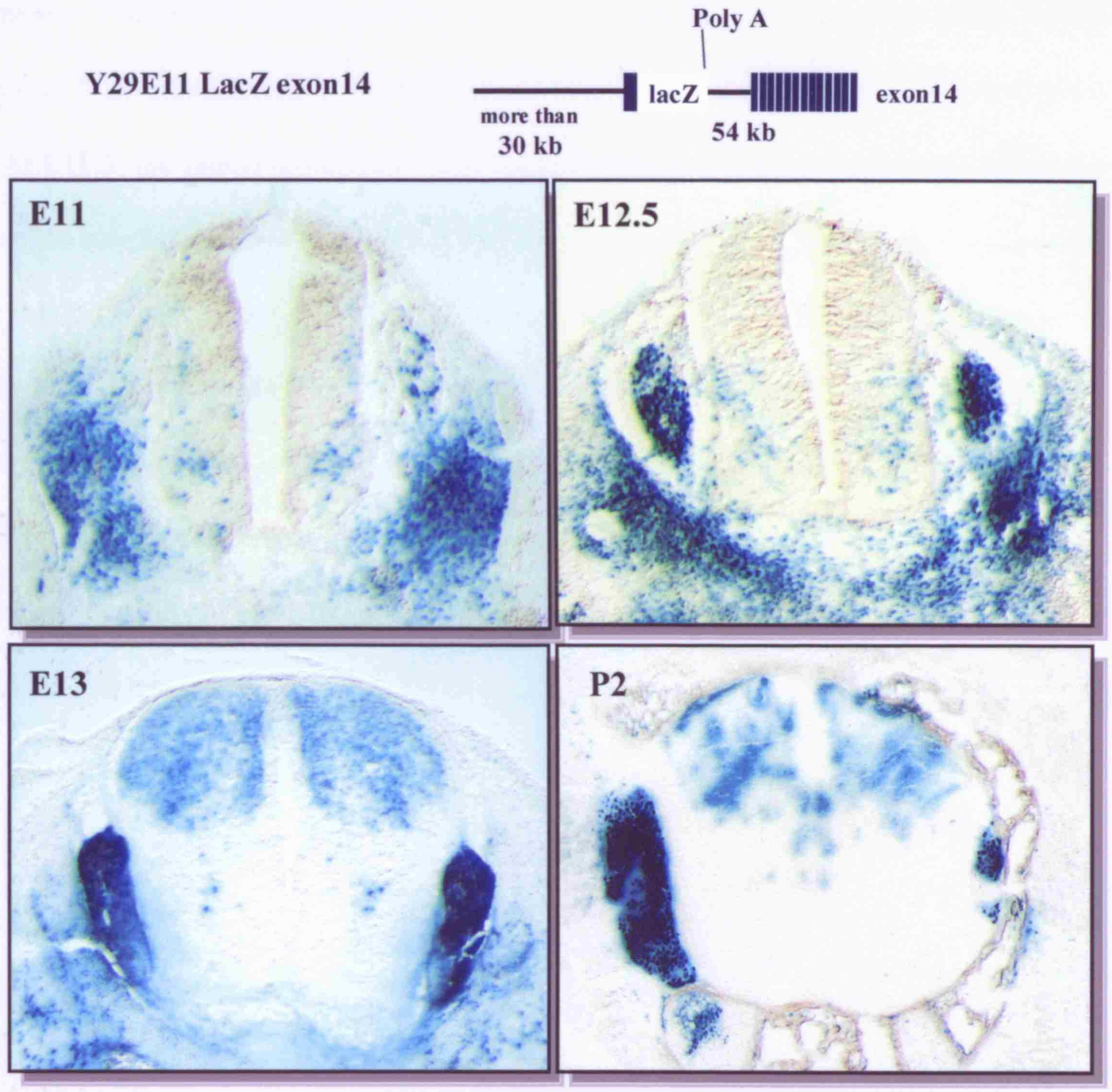

Y29E11 LacZ exon14
Poly A
lacZ
exon14
more than
30 kb
54 kb
E11
E12.5
E13
P2

Fig 3.12. Analysis of the embryonic spinal cord expressing the Y29E11LacZexon14 transgene. Transverse sections of E11, E12.5, E13 and P2 cervical spinal cords. There was strong β-gal staining at E11 in the sclerotome, neural crest cells and cells at the spinal cord. At E12.5, the DRGs are already well formed and are strongly labelled, as well as the sclerotome surrounding the vertebrae and cells outside the subventricular zone now only in the ventral spinal cord. Quick changes at the spinal cord morphology at E13 happened and the YAC transgene expression pattern also changes dramatically, most of the positive cells are in all the dorsal spinal cord but there is still a cluster of a few cells in the ventral part. The expression continues at the dorsal spinal cord at P2 but the cells seem mainly distributed in the grey matter. Chondrocytes in the vertebrae and DRGs kept also expressing this transgene until P2.

Y29E11Exon 2	E11	E12.5	E13-E15.5	Y449C2 (Sun,2000)	Endogenous m *Pdgfra*
Neural crest cells	+	+	+	+	+
Sclerotome	+	+	+	+	+
1st and 2nd Branchial Arches	+	+	+	+	+
Mesenchyme: Lungs	-	-	-	+	+
Nostrils	-	-	+	**Nd**	+
Liver	-	-	-	-	+
Tongue	-	-	+	+	+
Heart valves	-	-	-	-	+
Kidney	-	+	+	**Nd**	+
Stomach	-	-	-	**Nd**	+
Intestine	+	+	+	**Nd**	+
Skin	+	+	+	+	+
Dermal hair papillae (whiskers)	-	-	+	**Nd**	+

Table 3.1. The Y29E11LacZexon2 expression pattern at different embryonic stages and its comparison with the wild type mouse *Pdgfra* expression. E11, E12.5 and E13.5-15.5 embryos were carefully analysed. The category from E13.5 to E15.5 includes the results from the first transgenic embryo sacrificed at E15.5. The Y449C2 human *PDGFRA* transgenic and the endogenous mouse *Pdgfra* expression pattern are also included for comparison. Positive expression (+) or no expression (-) of the Y29E11LacZexon2 per cell type or organ is determined after the study of the β-gal staining and LacZ in situ hybridization in wholemount and/or cryostat sections. Nd, not determined.

Y29E11 Exon14	E11	E12.5	E13.5-E14	P2	Y449C2 (Sun 2000)	Endogenous *mPdgfra*
Neural crest cells	+	+	+	-	+	+
Sclerotome	+	+	+	+	+	+
1st and 2nd Branchial Arches	+	+	+	**Na**	+	+
Mesenchyme: Lungs	-	-	-	-	+	+
Nostrils	-	-	-	**Nd**	**Nd**	+
Liver	-	-	-	-	-	+
Tongue	-	-	+	**Nd**	+	+
Heart valves	-	-	-	**Nd**	-	+
Kidney	-	-	-	**Nd**	**Nd**	+
Stomach	-	-	-	-	**Nd**	+
Intestine	-	-	-	+	**Nd**	+
Skin	-	-	-	-	+	+
Dermal hair papillae (whiskers)	-	-	-	-	**Nd**	+

Table 3.2. Summary of the expression pattern of the Y29E11LacZexon14 transgenic line. The E11, E12.5, E13.5-14 and P2 embryonic stages were analysed. The P2 stage showed strong expression in the mesenchyme of the intestine, which was the only internal organ to show any expression of this transgene, no sections of the head at P2 were analysed. The Y449C2 human *PDGFRA* transgenic and the endogenous mouse *Pdgfra* expression pattern are also included for comparison. Positive expression (+) or no expression (-) of the Y29E11LacZexon14 per cell type is determined after careful analysis of the β-gal staining and LacZ in situ hybridization in wholemount and/or cryostat sections. Not determined (Nd) expression means that it was not analysed in this particular stage or transgenic line. Not applicable (Na) means that the tissue is not present anymore at P2 stage.

3.2.6. YAC transgene expression in pre-implantation embryos

In order to analyse the expression of Y29E11LacZexon2-B and Y29E11LacZexon14 transgenes in pre-implantation embryos, transgenic males were mated with wild type superovulated females. Approximately 120 embryos 0.5 day old were extracted from six superovulated females; they were incubated from 24 hrs to 72 hrs at 37°C and 5% CO_2 in M16 medium. After 24 hrs of incubation, many eggs were at four cell stage and morula stages. Later at 48 hrs and 72 hrs of incubations, many embryos were at morula, early blastocyst and blastocyst stages. The first detection of β-galactosidase expression of both YAC transgenes was at morula stage, and it continued in early blastocyst and blastocyst stages. The expression of both YAC transgenes was concentrated in four or five big cells at one side of the morula and in the early blastocyst already breaking the zona pellucida (Fig 3.13).

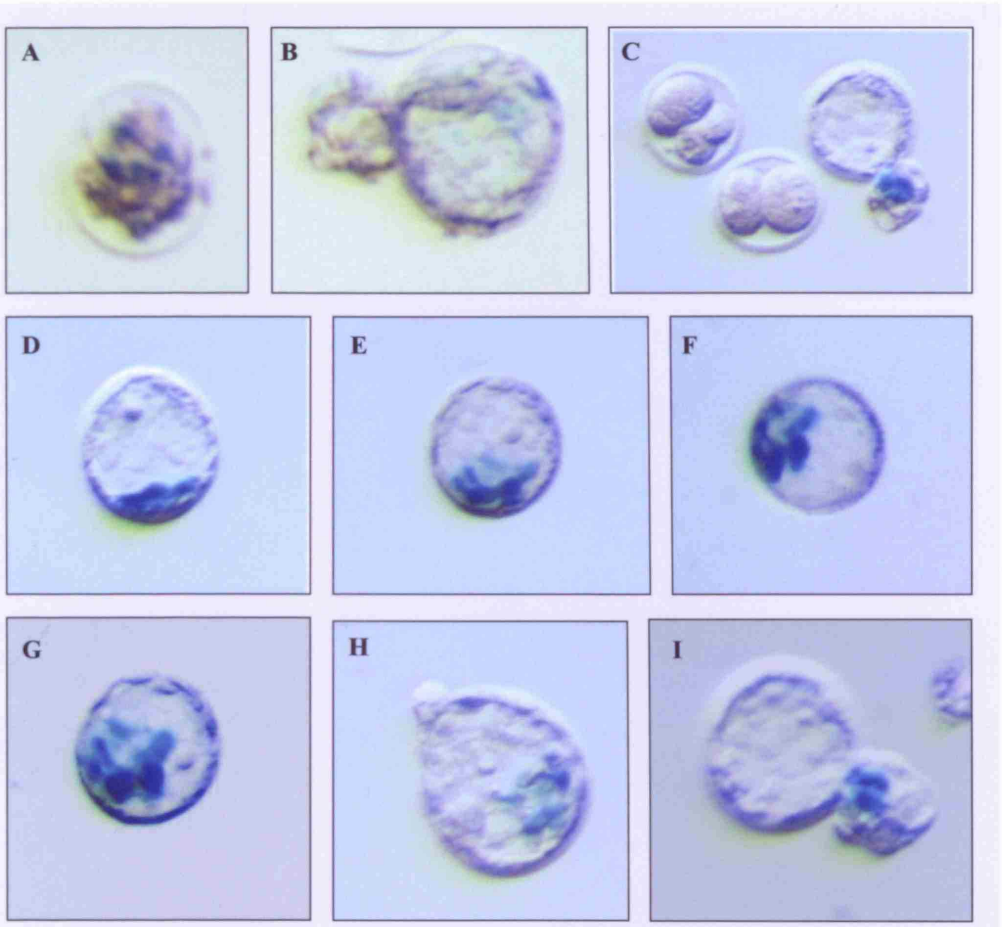


Fig 3.13. Pre-implantation transgenic embryos from the Y29E11LacZexon2 (A, B) and Y29E11LacZexon14 (C to I) mice transgenic lines. Wholemount β-gal staining. No staining was detected in two-cell stage embryos (C). However, embryos at morula stage showed around four cells strongly labelled at one side of the embryo (A, D,E,F,and G). Early blastocysts, already breaking the zona pellucida, have also cells that continue expressing the human *PDGFRA*-YAC transgenes (B, H and I).

3.3 Discussion

3.3.1. The Human *PDGFRA* YAC transgenes are first seen expressed at morula stage and later in mesoderm-derived tissues.

The very early embryo has three primary germ cell layers, endoderm, mesoderm and ectoderm. The mesoderm is the middle layer, which gives rise to the notochord, somites and mesenchyme. These three major structures and tissues differentiate to give rise to bone cells, cartilage, smooth muscle cells, the circulatory, lymphatic and excretory systems, and supportive and connective tissue. During development the mesoderm gets organised in two masses of cells alongside the neural tube called somitomeres, these compact and break up to create the somites. The somites are divided in three main groups of cells from top to bottom, the dermatome, myotome and sclerotome, which will give rise mainly to dermis, skeletal muscle and the vertebrae respectively. On the other hand, the mesenchyme, also called embryonic connective tissue, is loosely packed and contains collagen bundles, fibroblasts and unspecialized cells. Indeed, some of the mesodermal cells retain the capacity to differentiate in diverse directions; these multipotent cells are mesenchymal stem cells or marrow stromal cells. They can differentiate into osteoblasts, chondrocytes, myocytes, adipocytes, neuronal cells and beta-pancreatic islets cells.

PDGF receptors were found on the surface of mesenchymally-derived cell types and were associated with proliferation and chemoattraction of these cells (Ross et al., 1986). In 1988, the expression of growth factor transcripts in single or small numbers of

preimplantation mouse embryos was examined with the use of single-cell mRNA phenotyping (Rappolee et al., 1988). Among those growth factors analysed was *PdgfA*, which mRNA was found in all mouse preimplantation embryos till blastocyst stage. This expression of *PdgfA* and other growth factors in mouse blastocysts suggested that they play an important role in the growth and differentiation of early mammalian embryos. *PdgfA* mRNA was also found even in the unfertilized egg. This revealed *PdgfA* as a maternal transcript already in the ovulated oocyte, suggesting that the control of growth and differentiation during mammalian embryogenesis is regulated by growth factors from embryonic and/or maternal sources. Similarly, in very early Xenopus embryos two different forms of the *pdgfA* are shown to be encoded by maternal mRNAs (Mercola et al., 1988). Later using RNAse protection analysis, it was described that *PdgfA* and *Pdgfra* mRNAs are present during early post-implantation stages of mice embryos (Mercola et al., 1990). Further analysis of *PdgfA* and *Pdgfra* expression in pre-implantation and early post-implantation mouse embryos was carried out using in situ hybridisation. The results showed that at two-cell and blastocyst stages, all cells express both ligand and receptor mRNA and protein. In contrast, early post-implantation embryos express the ligand and the receptor but in different germ layers; *PdgfA* mRNA is expressed in embryonic ectoderm, while *Pdgfra* mRNA is localized in the mesoderm layers of embryonic and extra-embryonic membranes. Because of this observation, it was proposed that chronic autostimulation of Pdgfra occurs in pre-implantation embryos, whereas, following implantation, early mesoderm development is dependent on stimulation by ectodermally produced PdgfA (Palmieri et al., 1992). In summary, all these findings prompt speculation that PdgfA and Pdgfra promote the proliferation and migration of the early mesoderm.

Our human *PDGFRA*-YAC transgenes were expressed at early morula stage (Fig. 3.13); this shows that even though the egg could contain maternal *Pdgfra* mRNA, the early embryo starts activating the expression of the zygotic gene at least at morula stage. More importantly for our study, these results show that the regulatory elements to express this gene during early development are present in both YAC constructs. The most probable location of these early mesoderm-regulatory elements could be in the upstream sequence of the human *PDGFRA* gene. The Y29E11LacZexon2-B contained less than 30 kb upstream sequence so regulatory elements in that sequence seem enough to induce this early activation. Both human *PDGFRA*-YAC constructs were later expressed strongly in many mesenchymal and sclerotome derived cells, which will be discussed later.

No study on post-implantation embryos was done on any of the human *PDGFRA*-YAC transgenic lines. It would be interesting to see if these few *lacZ* positive cells are the ones that form the early mesoderm; if so, the specific movements and further development of the mesoderm would be easily studied in these transgenic mice models. In that way we could understand more the importance of PDGFA and PDGFRA during the proliferation and migration of the early mesoderm, and study the autocrine/paracrine mechanisms and relevance of this ligand/ receptor interactions during development.

3.3.2. Comparison between the expression patterns of the two human *PDGFRA* YAC transgenes and previous transgenic lines.

The comparisons between my human *PDGFRA*-YAC transgenics and previously described transgenics are relevant to my research mainly to estimate the location of specific regulatory elements relative to the human *PDGFRA* gene. All the previous human *PDGFRA* transgenes, including our YAC transgenes, are expressed in cranial neural crest cells, sclerotome and mesenchymal derivatives (Table 3.3).

Making direct comparison between the 2.2 kb-LacZ Human *PDGFRA* (Zhang et al., 1998) and our human YAC transgenic lines at E11, the three transgenes show *LacZ* expression in sclerotome and mesenchymal-derived cells, cranial neural crest cells and the first and second branchial arches (Fig.3.14). At E12.5, the three embryos express *LacZ* in a cluster of cells in the mesenchyme of the jaw articulation and in many sclerotome cells around the forming vertebrae and limb bone. Later in development at E13.5, some but not all of the neural crest, sclerotome and mesenchymal derivatives continue expressing the transgenes. For example, comparing the 2.2 kb-LacZ Human *PDGFRA* and Y29E11LacZexon2 transgenes, both were expressed in hair follicles of the whiskers, and mesenchyme of the nostrils and eye lids. The Y29E11LacZ exon14 will also be expressed in the mesenchyme of the eye lids but later at E14.5 (Fig. 10, E14.5). Comparing the 2.2 kb-LacZ Human *PDGFRA* and Y29E11LacZexon14 transgenic lines both had expression in the trigeminal ganglion from E11 to E13.5 (Fig 3.14); however, the Y29E11LacZexon2 transgene was never expressed in those cells.

Some cells seem to be expressing the human *PDGFRA* transgenes even though they do not show expression of the *Pdgfra* by in situ hybridisation (Table 3.3). This ectopic expression in cells like motoneurons, trigeminal ganglion cells, cells in the dorsal spinal cord VZ and in the DRGs, seems to be repeated in different human *PDGFRA* transgenic lines, showing that it could be mainly an effect of the DNA sequences introduced in the mouse genome, but not an artefact due to positional integration, type of vector or length of the construct.

All the previous human *PDGFRA* transgenes showed consistent strong expression in specific cell types showing that the YAC/ plasmid vectors and methods are not affecting the specificity of the human *PDGFRA* or related *LacZ* gene activation.

More importantly, this research shows that the *PDGFRA* gene expression is essential for the proper development of the cranial neural crest cells, sclerotome derivatives and some mesenchymal derivatives in mammals. *PDGFRA* could be helping these cells initially for proliferation, mainly for fast migration and perhaps even for differentiation at the right places in the embryo.

3.3.3 Expression of the human *PDGFRA* YAC transgenes in neural crest cells

The neural crest (NC) cells are a very migratory and versatile lineage distributed at the four axial levels of the embryo. The cranial NC cells are formed at the anterior neural tube

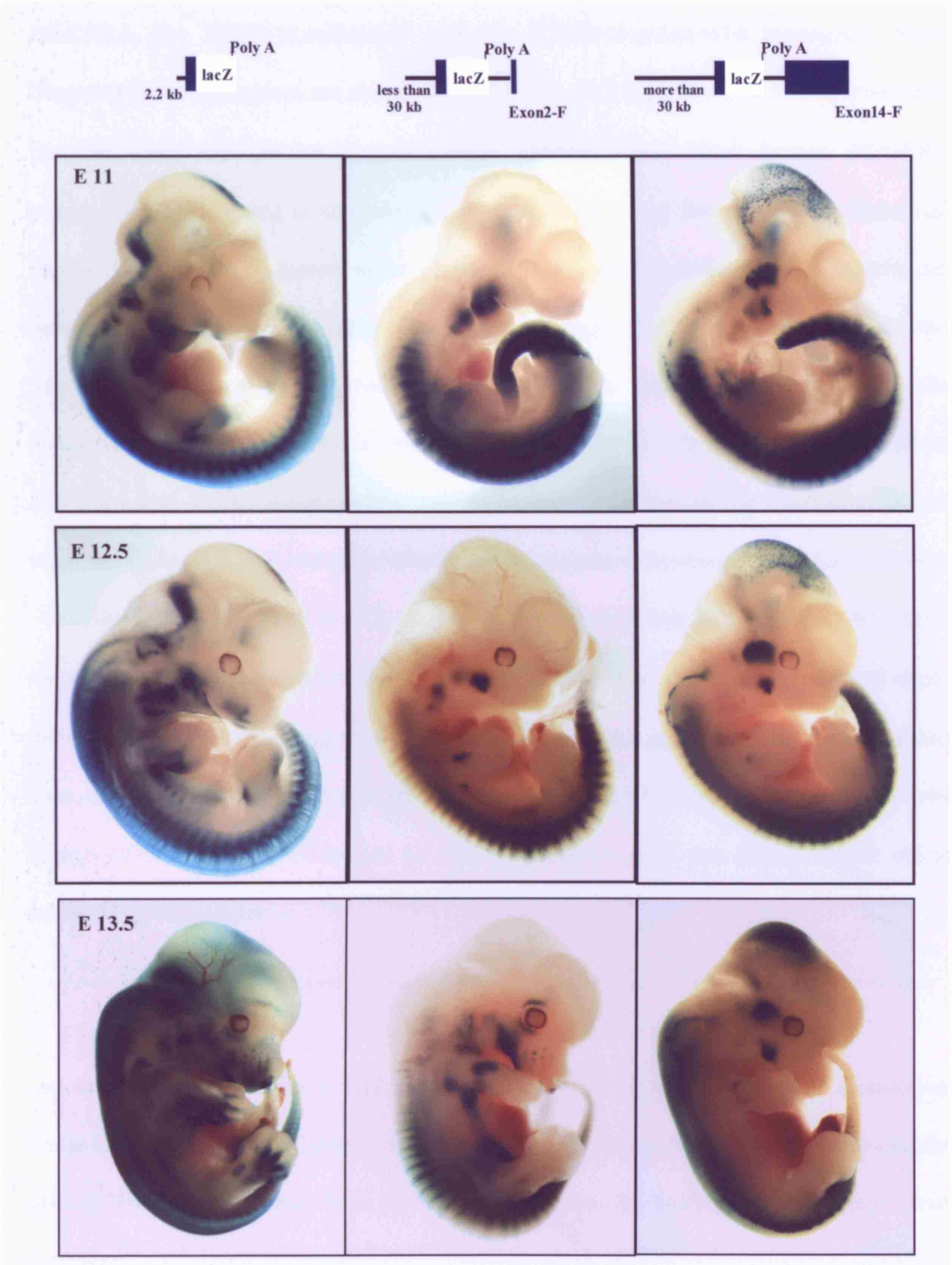

Poly A
lacZ
2.2 kb
Poly A
lacZ
less than
30 kb
Exon2-F
Poly A
lacZ
more than
30 kb
Exon14-F
E 11
E 12.5
E 13.5

Fig 3.14. Comparison between the expression pattern of the 2.2 kb-LacZ Human *PDGFRA*, the Y29E11LacZexon2 and the Y29E11LacZexon14 transgenic lines. Diagrams of the transgenes are shown above the E11, E12.5 and E13.5 embryos from each line, all were wholemount β-galactosidase stained. These three human *PDGFRA* transgenes are expressed in sclerotome, neural crest cells, and first and second branchial arches. Later in development some of the neural crest, sclerotome and mesenchymal derivatives continue expressing in the transgenic lines. For example, at E12.5 all the transgenes are expressed in a cluster of cells at the jaw articulation and in many cells during vertebrae and limb bone formation. In addition, comparing the 2.2 kb-LacZ Human *PDGFRA* and Y29E11LacZexon14 transgenic lines both had strong expression in the trigeminal ganglion and the mesenchyme of the tongue; comparing 2.2 kb-LacZ Human *PDGFRA* and Y29E11LacZexon2, they were expressed in hair follicles of the whiskers, mesenchyme of the nostrils and eye lids. The β-galactosidase staining method gives slight different labelling results due to the type of *LacZ* reporter gene cloned in the transgenic lines, in the 2.2 kb-LacZ Human *PDGFRA* the *LacZ* gene labels the entire cell, but in both human *PDGFRA* YAC transgenes an *nls-LacZ* reporter gene was cloned, which stains mainly the cell nucleus.

---

and they differentiate into a variety of cartilage, bone, cranial ganglia and connective tissue of the head and neck. Cardiac neural crest contributes to the vascular smooth muscle cells (VSMC) of the aortic arch. More caudal regions, the trunk and vagal neural crest contribute to the trunk peripheral nervous system and the enteric nervous system as well as melanocytes in the skin. The endogenous mouse *PDGFRA* gene seems to be very strong

and diffusely expressed in all the cranial neural crest derivatives. Similarly, our Human *PDGFRA* YAC transgenic founders showed expression in cranial neural crest cells and many of their derivatives. However, this human YAC transgene-cranial neural crest expression was in more restricted areas than the wild type *Pdgfra* expression. Other transgenes, the 6kb-LacZ human *PDGFRA*, 2.2 kb-LacZ human *PDGFRA* and Y449C2, were also expressed in cranial neural crest, but also only in areas near to the neural tube (Reinertsen et al., 1997;Zhang et al., 1998;Sun et al., 2000).

The first mouse model to show the importance of *Pdgfra* gene expression for cranial neural crest development was the *Patch* (*Pth*) mutant. As already mentioned, the deletion of the *Pdgfra* gene in the *Pth* mice results in early defects in the mesodermally-derived mesenchyme (Schatteman et al., 1992) and defects in neural crest derivatives. At E12 the *Pth* embryos showed a characteristic facial cleft -the maxillary process was of normal size, but the frontonasal and mandibular process were small and failed to fuse at the midline. Most embryos also had a cleft palate, a shortened neck and spina bifida beginning at the cervical level. In these *Pth* mice, the cranial neural crest cells that do not give rise to neurons (non-neuronal cranial neural crest) are particularly affected, which suggested that Pdgfra is required for the normal development of non-neuronal neural crest (Morrison-Graham et al., 1992).

Such drastic effects on the face and neck morphology due to the lack of Pdgfra correlates also with the fact that many muscles bone and cartilages of the face and neck develop from the first and second branchial arches, which have neural crest origin. Both of our human *PDGFRA* YAC transgenic lines show expression in the first and second branchial

arches. This confirmed the finding of Zhang (1998), the 2.2 kb-LacZ human *PDGFRA* transgene was also expressed in the mesenchyme of the first branchial arch at E8.5 and in the maxillary and mandibular processes at E12.5 (Fig 3.14). In addition, the 2.2 kb-LacZ human *PDGFRA* transgene was found in areas of mesenchyme in the upper and lower lips, eyelids, nostrils and pharynx. Similarly, the Y29E11LacZexon2 and the 6kb-lacZ *PDGFRA* transgene were expressed in mesenchyme of the eyelids and nostrils (Fig 3.6, D; Fig 3.14, Table 3.3).

The basal part of the heart is also of neural crest origin, and in keeping with this, the endogenous *Pdgfra* is expressed strongly in the basal part of the heart, and the aortic valve (Takakura et al., 1997;Sun et al., 2000). The important role of Pdgfra in cardiac neural crest cells is evident in the neural crest cell (NCC)-*Pdgfra* mutant embryos, which have *Pdgfra* conditionally deleted in neural crest cells. These embryos showed aortic arch defects because they are deficient in cardiac neural crest cells, but it is unknown whether this is due to the lack of proper function, differentiation or reduced number of these cells (Tallquist and Soriano, 2003;Hoch and Soriano, 2003). The 2.2kb-LacZ human *PDGFRA* was the only transgenic line that expressed in the heart, its signal was concentrated in the heart-valves and the walls of the large arteries at the base of the heart at E14.5 (Zhang et al., 1998). However, no evident expression of any of our human *PDGFRA* YAC transgenes, the 6kb-LacZ *PDGFRA* and the Y449C2 (Reinertsen et al., 1997;Sun et al., 2000) was seen expressed either in the basal part of the heart or its arteries in the embryonic stages analysed. The reason why only the smallest transgene of *PDGFRA* express in cardiac neural crest cells and the rest did not is not yet clear. At E11 in both human *PDGFRA* YAC transgenes many cells seem to be migrating ventrally to many

organs during development but no *PDGFRA* positive cells were detected in the developing heart. Therefore, in contrast to the strong expression in cranial neural crest, there was no apparent expression of the Human *PDGFRA* YAC transgenic founders in any cardiac, trunk or vagal neural crest cell derivatives.

Pdgfra mRNA was detected in E9 embryos in the dorsal portion of the neural tube during neural tube formation. These cells were most probably early neural crest cells that normally stop expressing *Pdgfra* once the neural tube closes completely after E9.5 (Schatteman et al., 1992). Weston and collaborators showed that cells coexpressing *E-cadherin* and *Pdgfra* are present in the non-neural epithelium of the neural folds at E8.5-E9, also before the neural tube closes. These findings have started a controversy about the origin and type of these *Pdgfra+* cells because of the similarity between mesodermally derived fibroblasts and ectomesenchyme cells, which are thought to be derivatives of the neural crest. I have already mentioned that in mice, the *Pdgfra* gene is expressed in both mesodermally derived mesenchymal cells and by neural crest derived cells. The observations by Weston raise the possibility that at least some *Pdgfra+* ectomesenchyme originates from the lateral non-neural domain of neural fold epithelium. In their interpretation, these cells seem to indicate a site of a delayed local delamination of mesenchyme similar to involution of mesoderm during gastrulation that they name metablast. Considering this, they proposed the hypothesis that neural crest and ectomesenchyme are developmentally distinct progenitor populations and that at least some ectomesenchyme is metablast-derived rather than neural crest-derived tissue (Weston et al., 2004). This hypothesis is still to be tested and approved by other scientists. Our human YAC *PDGFRA* transgenic mice might be useful in testing these ideas by

double in situ hybridization of the *LacZ*+ cells with other early neural crest and epithelial cell markers to identify them and even follow them through their migration.

3.3.4. Expression of the human *PDGFRA* YAC transgenes in somites, sclerotome progenitors and vertebrae development.

Coming from the ventral part of the somite, the sclerotome gives rise to the vertebrae. Sclerotome progenitor cells migrate ventrally around the spinal cord and towards the notochord, these cells join the sclerotome cells from the other side and form the vertebral body, later they move dorsally again to completely surround the spinal cord and form the vertebral arch. The complete molecular mechanisms of sclerotome induction, chondrogenesis and morphogenesis are still to be investigated in order to reveal the multiple steps of vertebra formation(Monsoro-Burq, 2005).

The endogenous *Pdgfra* gene is expressed strongly in somites (Soriano, 1997;Takakura et al., 1997) and sclerotome progenitor cells forming the vertebrae (Orr-Urtreger et al., 1992;Soriano, 1997). Strong expression in the somites was also observed in all the human *PDGFRA* transgenic lines. The 2.2 kb-LacZ human *PDGFRA* transgenic line had *lacZ* expression in somites as early as E8.5, our Y29E11LacZexon2 transgenic line also showed strong expression in somites at E10 (Fig. 3.8, B,C). At E12.5, once the somite has spread and divided into dermatome, myotome and sclerotome, the *lacZ* expression was mainly seen in the sclerotome of our human YAC transgenic lines (Fig.3.8, A,G; Fig 3.9; Fig 3.11,A,B,C), as well as Y449C2 and the 2.2 kb-LacZ human *PDGFRA* transgenic lines (Zhang et al., 1998;Sun, 1999;Sun et al., 2000). These lines also show less intense expression in the mesenchyme of the developing dermis which could have dermatome

origin or mesenchymal origin, but no expression was ever seen in the myotome (Fig 3.8, F and Fig 3.7, E12.5 and E13.5).

The importance of the *Pdgfra* expression during vertebrae development is extremely clear again in the *Pth* mutant mice, as the vertebrae are deformed and open (Schatteman et al., 1992). At E9.5 in these mice there are no sclerotome cells surrounding the notochord and ventral spinal cord as compared with the normal. Later, at E18, the vertebral bodies seem well developed but the sclerotome progenitor cells do not seem neither to migrate dorsally to completely surround the spinal cord or to fuse in the midline to form the vertebral arch. The sclerotome cells seem to migrate only laterally and create as a result, open flat vertebrae, a deformed spinal cord and two subepidermal blebs, one each side of the spinal cord which form probably as a result of the gap of tissue and failed vertebrae development.

The cells at the floor plate of the neural tube and the notochord produce Shh during development. Shh has been discovered to induce the cells of the sclerotome to form the vertebrae (Fan and Tessier-Lavigne, 1994;Fan et al., 1995;Buttitta et al., 2003), Shh signalling seems to require its transcription factors Gli2 and Gli3 for sclerotome induction but the complete molecular mechanisms involved are not yet established. After all this expression pattern analysis, it seems that as a result of this Shh induction the sclerotome progenitors surround the notochord and the ventral part of the spinal cord, starting to form the vertebral bodies. This first steps seems to occur in the *Pth* mutant mice, due to the expression of Shh by the notochord and floor plate, and the formation of the vertebral body. However, the sclerotome progenitors must continue migrating dorsally to form the

normal vertebrae; this step could have a stronger requirement for continuous *PDGFRA* expression, in order to keep many sclerotome cells migrating and proliferating dorsally.

This evidence and the strong and continuous expression of all the human YAC *PDGFRA* transgenes in the sclerotome until late in development, show that PDGFRA is essential for the proper migration and proliferation of the sclerotome progenitors and therefore the correct formation of the vertebrae during mammalian development.

Human *PDGFRA* transgenic lines	**Endogenous *mPdgfra***	**2.2 kb-LacZ (Zhang, 1998)**	**6 kb-LacZ (Reinertsen ,1997)**	**Y29E11 LacZ exon2**	**Y29E11 LacZ exon14**	**Y449C2 (Sun,2000)**
Neural crest cells	+	+	+	+	+	+
Sclerotome	+	+	+	+	+	+
1st and 2nd Branchial Arches	+	+	+	+	+	+
Mesenchyme: Lungs	+	+	+	-	-	+
Nostrils	+	+	+	+	-	Nd
Tongue	+	+	Nd	+	+	+
Heart valves	+	+	-	-	-	-
Kidney	+	+	+	+	-	Nd
Stomach	+	-	+	+	-	Nd
Intestine	+	-	+	+	+	Nd
Skin	+	+	+	+	-	+
Dermal hair papillae	+	+	Nd	+	-	Nd
Nervous system: Oligodendrocyte precursors	+	-	-	-	-	+
Motor neurons	-	+	-	+	-	-
Cells in dorsal spinal cord	-	+	-	-	+	+
DRGs	-	-	+	-	+	-
Trigeminal ganglia	-	+	-	-	+	Nd
Eye: Eye lids	+	+	+	+	+	-
Eye lens	+	+	-	-	-	+

Table 3.3. Expression pattern of human *PDGFRA* YAC transgenic lines. Y29E11LacZexon2 and Y29E11LacZ exon14 transgene expression patterns are summarized as well as the 2.2 kb LacZ human *PDGFRA*, 6kb human LacZ human

PDGFRA, Y449C transgenic and the endogenous mouse *PDGFRA* expression pattern, which were included for comparison. Positive expression (+) or no expression (-) of the transgenes per cell type is determined after analysis of the β-gal staining, or in situ hybridization of the LacZ gene, human *PDGFRA*-3'UTR, or ECD-*Pdgfra* in wholemount and/or cryostat sections. Not determined (Nd) means that it was not analysed in this particular stage or transgenic line.

3.3.5. The human *PDGFRA* YAC transgene expression in the mesenchyme of different organs during development.

Lungs

The endogenous mouse *Pdgfra* mRNA was detected in the mesenchyme or embryonic connective tissue of most developing internal organs by in situ hybridisation. The *Pth* mutant E18 embryo was described as an epidermal sac filled with organs due to the lack of mesenchyme even inside the organs (Schatteman et al., 1992). However, even though many transgenic lines show expression in the mesenchyme of some organs, none of them seem to express the human *PDGFRA* gene in all the mesenchymal cells like the endogenous gene. The 2.2 kb-LacZ human *PDGFRA* transgene seems to be expressed in more mesenchymal organs than any of the other human *PDGFRA* transgenes -in the mesenchyme of the lung, pancreatic primordium, the tracheal wall and urogenital sinus during organogenesis from E9.5 to E15.5. Particularly at E15.5, there was an abundance of β-galactosidase-positive cells present in the mesenchyme of the lung, its level of expression starting to decline one week postnatally and disappearing at around P20

(Zhang, 1998). The 6kb-LacZ human *PDGFRA* transgene and the Y449C2 seem to express also in the mesenchyme of the lung but the expression is not exactly in the same number or types of cells as the endogenous *PDGFRA* expression. In fact, the Y449C2 transgenic mice seem to rescue most of cranial and spinal deformities of the *PDGFRA* null mice, but postnatally these rescued mice had small lungs which were probably the cause of death. It seems that Y449C2 transgene was not successfully expressed in the smooth muscle cells of the lung (Sun et al., 2000). None of our human YAC transgenes had *lacZ* expression in the mesenchyme of the lungs at any embryonic stage as expected. In fact, the Y29E11LacZexon14 was only expressed in the mesenchyme of the intestine and only at P2.

Intestine and kidneys

PDGF signalling is also involved during organogenesis of the kidneys and intestines. PDGFRA is expressed in the mesenchymal cells and seems to help mainly in their migration and recruitment under the epithelium to induce folding, which produces villus formation in the intestine (Karlsson et al., 2000) and the invagination for hair follicles (Karlsson et al., 1999). On the other hand, PDGFB/PDGFRB signaling is involved during the formation of the glomeruli in the kidneys; it seems that it is not required for proliferation or survival but mainly for the migration of the mesenchymal progenitor cells to the glomerular space (Lindahl et al., 1998). The 6kb-LacZ human *PDGFRA* and the Y29E11LacZexon2 transgenes were strongly expressed in the mesenchyme of the kidneys (Fig. 3.8, D, E) and intestines (data not shown). However, Y29E11LacZexon14 was expressed only in the mesenchyme of the intestine until P4. In addition, the 2.2kb-LacZ

human *PDGFRA* transgenic line was expressed only in the kidney mesenchyme. These results suggest that there might be a group of activator and repressor regulatory elements specific for the regulation of *PDGFRA* in the mesenchyme of these organs. Possibly inside the 2.2kb upstream region there is an activator element that is responsible for maintaining expression of this gene in the mesenchyme of the kidney.

Bones

The sclerotome progenitors form the cartilage scaffold of the spinal column and ribs. For other bones, mesenchyme progenitors create the perichondrium/periosteum which is strongly labelled by the endogenous *Pdgfra* gene. Similarly in our human *PDGFRA* YAC transgene the perichondrium/periosteum of the long bones of the anterior and posterior limbs were β-galactosidase stained, the expression started in the early limb buds at E11 and persisted later at E13.5, mainly in the articulations (Fig 3.8, A, Fig. 3.14). The Y449C2, 6kb-LacZ human *PDGFRA* and 2.2kb-LacZ human *PDGFRA* transgenes were also expressed in developing limb buds and in the chondrifying bone primordia of limbs, vertebrae and ribs. The regulatory elements to control the expression in these perichondrium/periosteum progenitor cells must be within the 2.2 kb upstream sequence. The rest of the sequence in the human *PDGFRA* constructs does not seem to alter or modify the expression in these cells.

Notocord

The notochord is the only mesoderm derivative that does not show endogenous *Pdgfra* expression. None of the human *PDGFRA* transgenic lines showed expression in the notochord either.

Tongue

Both human *PDGFRA* YAC transgenes were strongly expressed in the mesenchyme of the tongue as well as the endogenous *Pdgfra* gene and all the other transgenic lines (Fig. 3.6, D; Fig. 3.11, B arrow).

Skin, nostrils, eye lids, hair follicles and developing whiskers.

The Y29E11LacZ exon2 transgene was expressed in the mesenchyme of the skin, hair follicles, whisker follicles, eye lids, nostrils particularly alongside of the nose cartilage. On the contrary, the Y29E11LacZexon14 transgene was only express in eye lids. However, PDGFRA is critical for the proper development of the skin; loose mesenchyme is seen beneath the epidermis which is strongly labelled by the endogenous *Pdgfra* gene in wild type mice but it is missing in *Pth* mutant mice (Orr-Urtreger and Lonai, 1992;Takakura et al., 1996). Other transgenic mice, like the 2.2 kb LacZ human *PDGFRA* also showed β-galactosidase staining in the condensed mesenchyme of the skin and hair follicles of the whiskers (Fig 3.14, E13.5 and Table 3.3). Therefore, it is possible that repressor binding sites are found in the sequence between exon2 and exon14 that inhibit expression of the transgene Y29E11LacZexon14 in the mesenchyme of the skin, nostrils and hair follicles. Before finding any repressor binding site in the sequence, it is still necessary to clarify the developmental origin of these cells, whether they originate from the mesenchyme or the cranial neural crest and then look for specific transcription factors related to each cell type.

PDGFA/PDGFRA and Shh signalling is very important for the correct development of the hair follicle. Shh is expressed in the epithelia of the developing whiskers, hair, and teeth of mouse embryos. In whisker and hair development, Shh is expressed in epithelial cells before condensation of the underlying mesenchymal cells, which express the *PDGFRA* gene (Takakura et al., 1996;Iseki et al., 1996). The results suggested that Shh is secreted by the epithelium as a signaling molecule in epithelio-mesenchymal interactions that direct terminal differentiation of the mesenchyme. Direct interactions between this Shh signaling cascade and the PDGFA/PDGFRA gene expression or signaling have not yet been proven. Nevertheless both molecular signaling pathways seem essential for the proper development of the hair follicles and whisker follicles as well as many other structures.

3.3.6. The human *PDGFRA* YAC transgene expression in oligodendrocyte progenitors

In the CNS, *Pdgfra* is expressed in oligodendrocyte progenitors (OLPs), which are first detected in mice at E12.5 in the ventricular zone of the pMN domain in the ventral spinal cord. These OLPs later proliferate and migrate throughout the spinal cord before differentiating into myelin-forming oligodendrocytes. Many *Pdgfra* expressing OLPs persist during adulthood. However, neither the 2.2kb lacZ human *PDGFRA* nor the 6kb lacZ human *PDGFRA* had previously shown expression in the OLPs. Therefore, no OLP-specific regulatory elements (OLP-SREs) are located in up to 6 kb upstream sequence.

Nevertheless, the Y449C2 transgene did activate the expression of the human *PDGFRA* gene in OLPs (Sun, 1999;Sun et al., 2000); this YAC contains 3 kb upstream sequence, the complete structural gene and around 310 kb downstream sequence. Therefore, it was possible that the OLP-specific regulatory element(s) (OLP-SREs) was in the long 29 kb intron 1 or within the intron-exon structure of the gene or even anywhere in its 310 kb downstream sequence, which gave in total a sequence of around 376 kb to analyse. My study of the Y29E11LacZexon2 and Y29E11LacZexon14 transgenes confirmed that the OLP-SRE(s) is not in the up to 60 kb upstream sequence or within the intron-exon structure of the gene up to at least exon 14. Therefore, these OLP-SREs could be anywhere from intron 14-15 to 310 kb downstream. In Chapter 4 of this thesis, I describe the findings from another two human *PDGFRA* BAC transgenes that have helped to narrow down the region where the OLP-SREs might reside.

3.3.7. Ectopic expression of the *PDGFRA* transgenes in dorsal spinal cord cells, trigeminal ganglia, DRGs and motoneurons.

Some transient *Pdgfra+* labelling at the E13.5 dorsal spinal cord of the normal rat was seen in many cells distributed at the alar plate in most of the dorsal spinal cord. These cells are possibly developing interneurons. This expression seems to stop before E16 according to in situ hybridisation (Pringle and Richardson, 1993). Similar expression has been detected in transgenic mice like the 2.2kb-LacZ human *PDGFRA* and the Y449C2, which showed expression in cells of the dorsal spinal cord, their distribution looks similar but not identical and the beginning of the expression does not seem to be exactly at the same developmental stage in each transgenic line but it is maintained until P2. In addition, my

Y29E11LacZexon14 transgene was also expressed strongly in many cells at the dorsal spinal cord from E13, and it continued until P2 in cells mainly outside the VZ (Fig 3.12). These strongly-labelled dorsal spinal cord cells in the transgenic lines appear to be undifferentiated early precursors, which could later differentiate into dorsally originated interneurons, OLPs and astrocytes.

Motoneurons, DRGs and trigeminal ganglia.

The dorsal root ganglia (DRGs) are groups of neuronal cell bodies that lie along the vertebral column and bring sensory information to the brain. The DRGs develop in the embryo from neural crest cells. The ligand *PDGFA* mRNA is expressed in developing dorsal root ganglion (DRG) neurons beginning between E12 and E15 (Yeh et al., 1991). The effect of PDGFA on the number of DRG neurons begins at E12, when the DRG are first formed. In addition, PDGFAB has been reported to promote the survival of chick spinal motoneurons during the period of naturally occurring cell death (Oppenheim et al., 1993). *PDGFRA* transcripts were not detected in E13 mouse DRGs by in situ hybridization (Morrison-Graham et al., 1992). Nevertheless, a study in neonatal rat found satellite cells and possibly sciatic nerve neurons and DRG expressing *Pdgfra* (Eccleston et al., 1993). A quantification of the motoneurons and DRG neurons in the *Pth* mutant mice showed that the number and size of lumbar spinal motoneurons and DRG neurons are significantly decreased at E14 and E17 stages. Therefore, the lack of Pdgfra seems to affect the survival and morphological differentiation of spinal motor neurons and DRG neurons. Pdgfra might play a key role in the development of both motoneurons and DRG neurons. All these findings, suggested that neuronal produced PdgfA may act through

Pdgfra as a paracrine regulator in the CNS and as a paracrine and/or autocrine factor in the PNS during development (Li et al., 1996).

Another way that the lack of Pdgfra could be affecting DRG development is through the aberrant migration pattern of neural crest cells in *Pth* embryos during earlier development. DRG neurons originate from neural crest cells and their lack of migration could cause the DRGs to develop with fewer neurons. Therefore, it is the lack of Pdgfra earlier during development which could provoke the observed changes in the spinal cord and DRGs morphology.

The human *PDGFRA* transgenic lines have also shown some relation between *PDGFRA* expression and both motoneuron and DRG development. Some subsets of motoneurons seem to be strongly labelled in the basal plate of the ventral spinal cord in the 2.2 kb-LacZ human *PDGFRA* transgenic line and in both the Y29E11LacZexon2 and Y29E11LacZexon23 transgenic lines (Fig 3.9, Fig.3.12). The first founder of Y29E11LacZexon2 shows strong expression in motoneurons at E15.5 (Fig 3.6, B, C). However, the expression does not begin at the same stage in every founder. In case of the DRGs, the 6kb-LacZ human *PDGFRA* and the YAC29E11LacZexon14 transgene were expressed particularly strongly in the DRGs (Fig 3.11; Fig 3.12); other founders of other human *PDGFRA* transgenes have also shown expression but have not been reproducible. This motoneuron and DRG expression does not seem to show any direct relation to a specific fragment of DNA sequence inserted in the human *PDGFRA* constructs. A possible explanation for the DRG neuron expression could be their neural crest origin. The neural crest cells were strongly labelled in must of the human *PDGFRA* transgenic lines,

and these neural crest-derived cells might continue expressing the transgene when it should normally have stopped under the normal transcriptional regulation of the gene. It is important to note that mRNA stability could well be different among the different transgenes and this could cause the mRNA to persist longer in some cell lineages.

The cephalic ganglion cells of the trigeminal ganglion are also of neural crest origin. Mouse *Pdgfra* is not expressed in the trigeminal ganglion, but the Y29E11LacZexon14 as well as the 2.2 kb-LacZ human *PDGFRA* transgenic lines showed strong lacZ staining in many cells of the trigeminal ganglion (Fig 3.11, F; Fig 3.14).

In addition, as already mentioned the secreted signaling molecule sonic hedgehog (Shh) functions in sclerotome induction, limb pattern formation, and differentiation of motor neurons (Iseki et al., 1996). However, the way the Shh pathway induces differentiation of motoneurons is still being studied, and it is still to be established if this could activate indirectly or have any influence in the PDGFA/ PDGFRA signaling pathway. The complete study of these pathways is very important to understand how motoneurons develop normally in the embryo, and then fully reveal why the disruption of such pathways affects them so dramatically.

3.3.8. Expression of the human *PDGFRA* transgenes in the brain.

In the normal mouse brain, the first transcripts of *Pdgfra* were found in the roof of the fourth ventricle of E9.5 embryos. Later during development, the posterior choroid plexus, which develops from the fourth ventricle, and the anterior choroid plexus continue

expressing the *Pdgfra* gene throughout development. The meninges also express *Pdgfra* (Schatteman et al., 1992). Furthermore, neurons and glial cells in adult mouse parasympathetic ganglia seem also to express Pdgfra protein (Schatteman and M. Byers, unpublished results). Although the brain of *Pth* mutant embryos were not extensively examined, these mice had evident brain abnormalities; the embryos seemed to show defects in the choroid plexus as expected but each embryo had a variety of other problems, for example one had a small metencephalon, another did not have olfactory bulb and another had collapsed ventricles (Schatteman et al., 1992).

Cells in the hypothalamus and choroid plexus were positively labelled in both the human YAC transgenic mice (Fig. 3.6, G, H; Fig. 3.8, I; Fig 3.11, H, I). The Y29E11LacZexon2 was also expressing in a few cells in the corpus striatum mediale (Fig. 3.8, H arrow). In addition, the Y29E11LacZexon14 showed β-gal labelled cells all around the most external layer of the midbrain, a cluster of cells inside the midbrain, cells in the anterior pons, medulla oblongata and cerebellum (Fig 3.11, G-J). In this transgenic line, the midbrain and dorsal spinal cord are still showing strong expression in the P8 brain. The identity of these cells was not studied further.

The 2.2kb-LacZ human *PDGFRA* transgene labelled positively many cells in different areas of the brain (Fig 3.14), the number of cells and areas also seem to increase towards development, however none of them were endogenous mouse *Pdgfra* positive cells. There was not expression in the roof of the fourth ventricle, the choroid plexus, OLPs or meninges (Zhang et al., 1998). My own human YAC transgenic mice have less expression in most areas of the brain, they had correct expression in one area, the choroid plexus. It

seems that probably the extra upstream and up to exon 14 sequence of the transgenes help to improve the proper regulation of the human *PDGFRA* gene by repressing inappropriate expression in many areas of the brain.

3.3.9. Tentative localization of the human *PDGFRA* specific regulatory elements according to the transgenic analysis.

As mentioned by Zhang and Reinertsen, 6kb, 2.2 kb and even 0.9 kb-LacZ *PDGFRA* transgenes could contain critical specific regulatory elements to activate the expression of the *PDGFRA* gene (Reinertsen et al., 1997;Zhang et al., 1998). Similarly to those transgenic lines, our two Y29E11LacZexon2 and Y29E11LacZexon14 transgenes, which contained from 30kb to 60 kb upstream sequence, were also showing strong expression in specific cell types. There is still a small possibility that integration site-specific effects could confound my analysis, since I examined only a small number of YAC transgenic founders. However in my research, I consistently observed cranial neural crest cells, sclerotome progenitor cells, and mesenchyme derived cells strongly expressing the transgenes as Zhang et al. (1998) and Reinertsen et al. (1997). Therefore, I concluded that PDGFRA is essential for the proper development of cranial neural crest cells, sclerotome progenitors and mesenchyme derived cells, in particular for their adequate proliferation and migration in the embryo.

In conclusion, I confirm the findings by Zhang et al (1998), indicating that the 2.2kb upstream sequence could be enough to activate the specific *PDGFRA* gene expression in cranial neural crest cells, sclerotome progenitor cells and mesenchyme derived cells.

These new human YAC transgenic mice showed that the extra upstream sequence and intron 1 did not seem to influence dramatically the *PDGFRA* expression. In any case, the extra upstream and downstream sequence as far as exon14 could contain regulatory sequences for repressors/inhibitors of expression, as suggested by the lack of the human *PDGFRA* expression in many cell types in the Y29E11LacZexon14 transgenic line in comparison with the other two smaller constructs (table 3.3 and Fig 3.14). Repressor regulatory elements are known to be important for the correct control of gene transcription during development. More discussion of the location of SREs that direct expression in cranial neural crest cells, sclerotome progenitor cells, and mesenchyme derived cells will be included in Chapter 4.

Chapter 4

OLP-regulatory element(s) at the 3'end of the Human *PDGFRA* gene sequence.

4.1 Introduction

4.1.1 Oligodendrocyte precursors and the debate about their single or multiple sites of origin in the neural tube.

Oligodendrocyte progenitor cells (OLPs) were first identified *in vitro*, in cultures of perinatal rat optic nerve cells, and a bipotential glial cell precursor named 02A was described that could give rise to type-2 astrocytes and oligodendrocytes (OLs) *in vitro* (Raff et al., 1983;Raff, 1989). These O2A progenitors are bipolar in shape, respond to PDGF by mitosis and express the *Pdgfra* gene (Hart et al., 1989). However, the idea of a common precursor for oligodendrocytes and astrocytes has been put into question by the discovery of different sites of origin for OLPs and astrocyte precursors in the ventricular zone (VZ) of the neural tube (Pringle and Richardson, 1993;Yu et al., 1994;Pringle et al., 1998;Pringle et al., 2003).

The establishment of the specific site or sites of origin of OLPs in the developing CNS has involved years of controversy and debate between different research groups. Some groups suggested that OLPs originate only in a small region of the VZ in the ventral part in the spinal cord, named the pMN domain, from where they later proliferate and migrate to populate the entire spinal cord (Fig. 1.2) (Warf et al., 1991;Pringle et al., 1992;Pringle

and Richardson, 1993;Noll and Miller, 1993;Yu et al., 1994;Pringle et al., 1996;Hall et al., 1996;Pringle et al., 1998;Richardson, 2001;Kessaris et al., 2001;Pringle et al., 2003). Warf et al. (1991) studied the capacity of specific regions of embryonic rat spinal cord to give rise to OLs *in vitro*. At E14, they detected that all ventral regions along the rostral-caudal axis of the spinal cord have the capacity for oligodendrogenesis. By contrast, dorsal regions of the thoracic and lumbar spinal cord do not develop the capacity for oligodendrogenesis until later in development. Based on these results, they proposed that the capacity of dorsal rat spinal cord to give rise to OLs appeared to be associated with ventral-to-dorsal migration of OLPs (Warf et al., 1991). Later, the *Pdgfra* expression pattern in embryonic rat brain (Pringle et al., 1992) and spinal cord (Pringle and Richardson, 1993) was analysed and detected in presumptive OLPs but not in neurons. First, it was shown that many scattered *Pdgfra+* cells are distributed throughout the rat brain at E16 (Pringle et al., 1992). In the spinal cord, small numbers of *Pdgfra+* cells first apper in the VZ of the ventral spinal cord at E14, and later these cells seem to proliferate and migrate away from the VZ and become widely distributed in the spinal cord even at the most dorsal regions (Pringle and Richardson, 1993). Since then the *Pdgfra* gene was established as a specific marker of OLPs *in vivo*, which seemed to be generated only in the ventral part of the spinal cord (Pringle and Richardson, 1993). Other workers (Yeh et al., 1993) showed that Pdgf-A is expressed by neurons in the adult and embryonic spinal cord and brain, suggesting that paracrine interaction between neurons and glial progenitors is responsible for OLP development.

Supporting the idea of a single ventral OLP origin, Noll and Miller also noted that OLPs were initially restricted to ventral regions of the embryonic rat spinal cord after analysing

cells at different stages of OL development. They used the markers A2B5 to label OLPs, 04 for immature and anti-galactocerebroside (GC) for mature OLs. In addition, they did BrdU labeling of these cells and mentioned that OLPs or their progeny migrate dorsally and laterally during subsequent spinal cord development (Noll and Miller, 1993).

A new suggestion considering multiple sites of OLP origin in the VZ of the spinal cord was made by Cameron-Curry and Le Douarin, after the analysis of 19-26 somite stage quail-chick chimeras. Labeling with the quail-specific marker SMP (Schwann cell myelin protein, a glycoprotein characterized as a precocious marker for Schwann cells and OL) in the chimeric spinal cords suggested that OLs were generated in the VZ of not only the ventral but also the dorsal part of the spinal cord. It seemed that most of the VZ was able to generate OLs except the regions closer to the roof and floor plate. In addition, by careful observation of the quail or chick cells migrating from their grafts, they also suggested that the OLPs migrate in both dorsoventral and ventrodorsal directions (Cameron-Curry and Le Douarin, 1995).

In the meantime, using immunoselection and culture of Pdgfra+ rat spinal cord and optic nerve cells, it was shown that the Pdgfra+ cells differentiated into OLs and co-labeled with A2B5 and O4 markers. Only few OLs develop in culture when the Pdgfra+ cells were removed. In addition, it was shown that E14 ventral rat spinal cord cells generate OLs but dorsal cells only acquire this ability after E16. These data demonstrate that most or all early-developing OLs are derived from the ventrally-derived Pdgfra+ OLPs (Hall et al., 1996).

Subsequently, the chick-quail chimeras were repeated, in which dorsal or ventral segments of the chick neural tube were replaced with equivalent segments of quail grafts and labelled with SMP. However, in this study only ventral grafts gave rise to OLs but not dorsal grafts (Pringle et al., 1998). These different results from similar techniques added to the controversy, and maintain the idea that OLPs were only ventrally originated from the pMN domain in the spinal cord. In addition, the analysis of heterozygous Danforth's short tail (Sd+) mutant mice, which lack caudal fragments of notochord and floor plate, showed that OLPs did not appear at the ventral VZ when there was no floor plate. Explants of this ventral VZ did not develop OLPs *in vitro*, which seem to suggest that the Shh signal from the notochord and floor plate are necessary to induce the development of ventrally derived OLPs (Pringle et al., 1996). These results support the ventral-only origin of OLPs in the spinal cord. A similar unique ventral OLP-origin and strong dependence on Shh was also suggested in the telencephalon (Tekki-Kessaris et al., 2001).

Around the same time, the *PDGFRA* expression pattern analysis in human spinal cords was made by Hajihosseini et al. (1996), discovering two clusters of two to three OLPs that appear at 45 days post-conception (dpc), or possibly before, on each side of the ventral cord ependyma. These OLPs were labeled with O4, PDGFRA, myelin basic protein (MBP), 2',3'-cyclic nucleotide 3' phosphodiesterase (CNP) and proteolipid protein (PLP) in ventral and lateral regions of the spinal cord. Later at 74 dpc and at 83 dpc, OLPs appeared at the dorsal spinal cord regions. By observing the dispersion pattern of these ventrally derived human OLPs, they also suggested that these cells progressively populate lateral and dorsal regions (Hajihosseini et al., 1996).

Alongside, a third proposal has been suggested considering that OLPs have multiple origins in the spinal cord and telencephalon, mainly because different OLPs expressed different markers, like the proteolipid protein (PLP/DM20) and the PDGFRA. Such markers are not co-expressed in the same cells in the spinal cord or brain of E10.5, E12.5 and P1 mice (Timsit et al., 1992;Timsit et al., 1995;Spassky et al., 1998;Spassky et al., 2000). However, it is possible that these markers are expressed at different stages of OLP development, rather than in different OLP lineages. Markers Olig1, Olig2, Sox10, A2B5 and PDGFRA are all expressed at earlier stages than SMP, PLP/DM20, O4, GC, MBP, P0 and CNP. Different groups also have worked with different animal models (rat, chicken, mouse and human) through the years which have added more data, but also discussion and disagreements.

After more than ten years of scientific debate about the precise site of origin of the OLPs in the VZ, the data from three papers published in 2005 helped to arrive at a consensus (Fogarty et al., 2005;Vallstedt et al., 2005;Cai et al., 2005). Two of these studies (Vallstedt et al., 2005;Cai et al., 2005) showed that apart from the ventrally derived OLPs, there were also $Olig2^{+}$, $Sox10^{+}$ and $PDGFRA^{+}$ OLPs in the VZ of the wild type dorsal spinal cord starting from E15.5 to P0. They described that the Nkx6.1/6.2 $^{-/-}$ mutant mice do not have ventrally derived OLs in their spinal cords, but the dorsally originated-OLPs nevertheless appear at E15 as normal. Therefore, these dorsally originating OLPs seem to generate all the OLs in the spinal cord of the Nkx6.1/6.2$^{-/-}$ mice. In conclusion, their data showed that OLPs are generated from both dorsal and ventral progenitor domains (Vallstedt et al., 2005;Cai et al., 2005). At the same time, experiments using Cre mediated recombination in transgenic mice to follow the cell fates of neural progenitors from the Dbx1/2-

expressing spinal cord VZ (two dorsal and two ventral domains; dP5, dP6, p0 and p1) showed that in addition to the expected interneuron populations, both OLs and astrocytes are generated from the Dbx1/2 domain (Fogarty et al., 2005). These Dbx-derived OLs comprise less than 20% of the total OL population in the cord and they mainly settle in the lateral white matter, opposite their point of origin. However, it showed that OLs could be generated from other domains in the spinal cord apart from the ventral pMN domain (Fogarty et al., 2005).

Overall the evidence shows now that OLPs are not only derived from the ventral pMN domain but also from the dorsal neuroepithelium of the spinal cord (Warf et al., 1991;Pringle et al., 1992;Pringle and Richardson, 1993;Noll and Miller, 1993;Yu et al., 1994;Cameron-Curry and Le Douarin, 1995;Pringle et al., 1996;Hajihosseini et al., 1996;Hall et al., 1996;Richardson et al., 1997;Spassky et al., 1998;Pringle et al., 1998;Spassky et al., 2000;Lu et al., 2000;Zhou et al., 2000;Richardson, 2001;Fogarty et al., 2005;Vallstedt et al., 2005;Cai et al., 2005;Richardson et al., 2006) and telencephalon (Grove et al., 1993;Spassky et al., 2001;Woodruff et al., 2001;Tekki-Kessaris et al., 2001;Kessaris et al., 2006). Postnatal waves of OLP production have also been detected in the telencephalon (Levison and Goldman, 1993;Levison et al., 1993), and in fact it seems that one wave of OLPs produced during early development is eliminated later during postnatal life and replaced or compensated by other waves of OLP production (Kessaris et al., 2006).

Discoveries during the OLP site of origin research have been essential for the correct interpretation of the human *PDGFRA* transgenic mice results concerning its expression in

OLPs. Of the previously made human *PDGFRA* transgenic mice, only the Y449C2 mice showed expression in OLPs (Sun et al., 2000). In this chapter, I show that *PDGFRA* is expressed in postnatal OLPs in human *PDGFRA* BAC transgenic lines, and narrow down considerably the genomic sequence region where the OLP *cis*-regulatory elements could be located. In addition, these results suggest a distinction between the regulatory mechanisms that activate the expression of *PDGFRA* in embryonic versus postnatal OLPs.

4.2 Results

4.2.1. Description of BACs containing the Human *PDGFRA* gene.

At the beginning of this PhD project, PAC and BAC technology was not well established. No BACs containing the complete human *PDGFRA* gene with upstream and downstream sequences were known. However, the human PAC library RPCII (High density gridded filters, UK HGMP Resource Centre, Cambridge) was already available, and I screened it with the human *PDGFRA* exon1 probe (PCR product) and the 3'UTR-human *PDGFRA* probe to find any PAC clones containing the *PDGFRA* gene. However, lots of non-specific positive clones were detected. Many of these clones were empty clones with a pUC-link fragment, which seems to induce the non-specific binding of the probes.

Through the years, human BAC clones that contained parts of the human *PDGFRA* gene had been described in the GenBank database, as shown in figure 4.1. The human BAC clone RP11-601I15 contains 97 872 kb upstream sequence and up to exon 2 of the human

PDGFRA gene, this BAC clone is approximately 130 kb long. Human BAC RP11-231C18 is around 132.1 kb long and its sequence starts from exon 3. Clone RP11-626H4 contains 3kb upstream sequence, the entire human *PDGFRA* structural gene and approximately 110 kb downstream sequence. This BAC clone was chosen for further analysis by transgenesis.

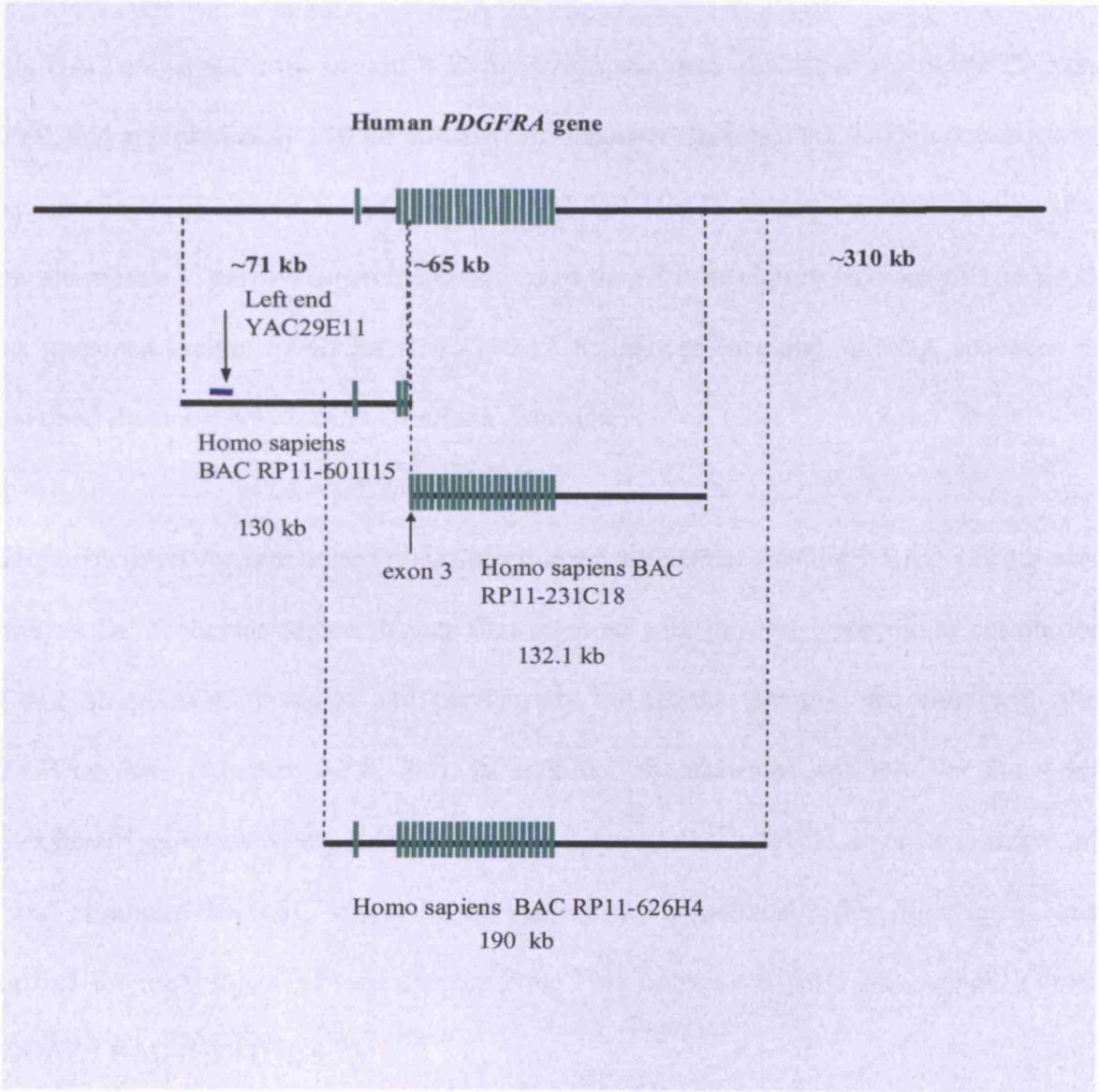


Fig 4.1. Schematic representation of Human *PDGFRA* BAC clones. Homo sapiens BAC clone RP11-601I15 is 130 kb long and contains 97 872 bp upstream sequence and up to exon 3 of the human *PDGFRA* gene. Homo sapiens BAC clone RP11-231C18 is 132.1kb long, its sequence starts from exon3. However, Homo sapiens BAC clone RP11-626H4 contains the complete human *PDGFRA* gene, 3kb upstream and 110 kb downstream sequence, it was later named human *PDGFRA*-BAC-110 kb. Exons are represented as green boxes and the black line is the non-coding sequence.

4.2.2. Purification and deletion of the human *PDGFRA* BAC constructs.

The human BAC clone RP11-626H4 was obtained from the BACPAC Resource Center (BPRC) at Children's Hospital Oakland Research Institute in Oakland, California, USA. This BAC contained only around 3 kb upstream sequence, similar to the Y449C2 (Sun 2000), and approximately 110 kb downstream sequence, less than the ~300kb downstream sequence in Y29E11 and Y449C2 (Fig. 4.2, 3.1, 4.11). Therefore, this BAC provided a new alternative to narrow down the locations of the OLP-regulatory element(s). The BAC was re-named human *PDGFRA* BAC-110 kb for convenience and its DNA sequence is contained in contig NT-022853 GenBank database.

The purification for pronuclear microinjection of the human *PDGFRA* BAC-110 kb was done by Dr. Francoise Jamen. It was first digested with the Not I enzyme to cut out its 11.612 kb pBACe3.6 vector and purify only the human genomic sequence with the *PDGFRA* gene (Chapter 2.2.8, 2.5). In addition, she truncated this BAC at the 47kb downstream sequence by digesting it with the enzymes AscI and ClaI, in order to fragment it and eliminate the BAC vector at the same time. Immediately after digestion it was purified for microinjection (see chapter 2.5). This fragmented BAC was named human *PDGFRA* BAC-47kb (Fig 4.2).

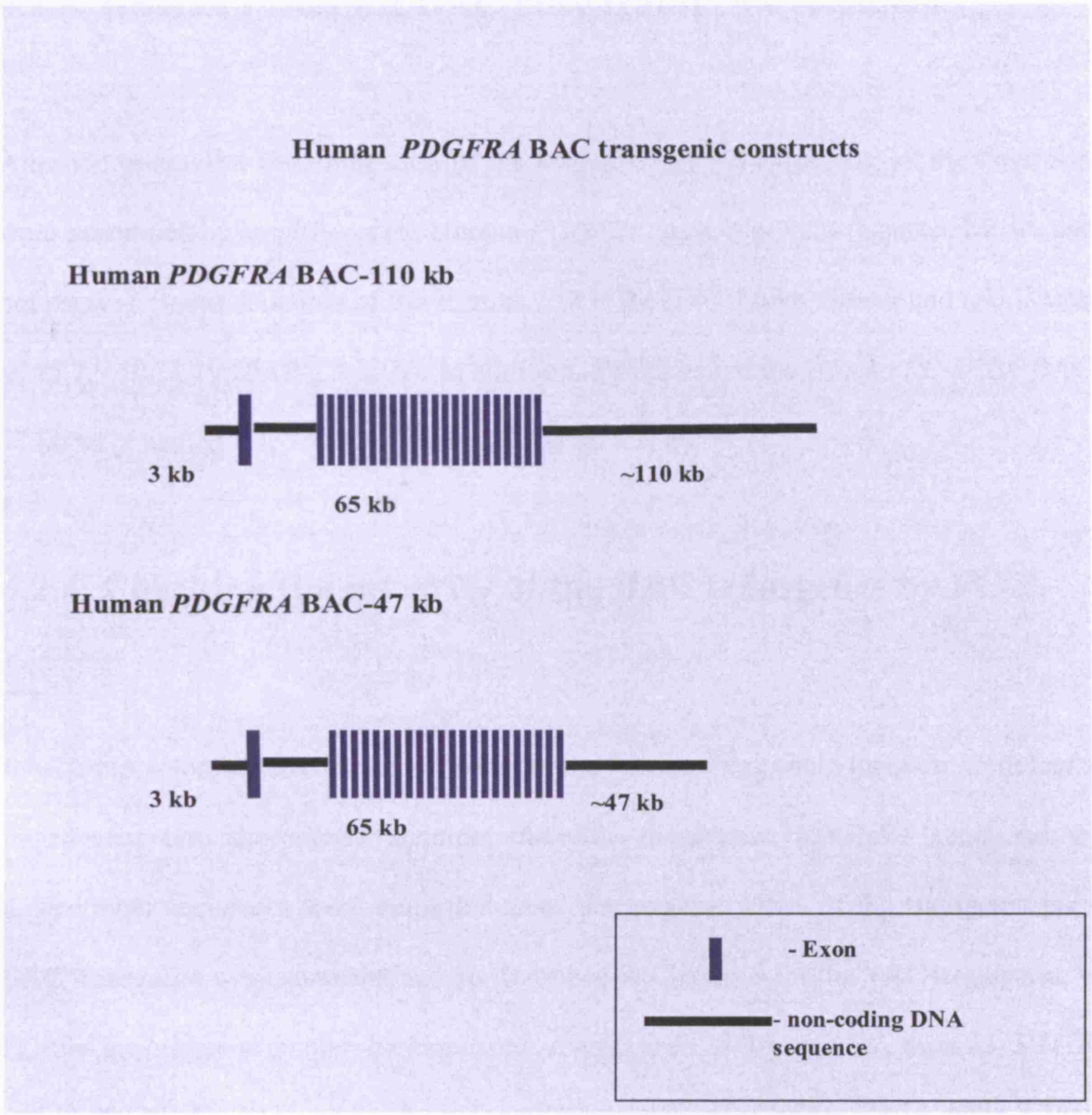


Fig 4.2. Schematic representation of the Human *PDGFRA* BAC 110 kb and fragmented construct Human *PDGFRA* BAC-47kb. Human BAC clone RP11-626H4 has the complete human *PDGFRA* gene, 3kb upstream and 110 kb downstream sequence, it was named human *PDGFRA*-BAC-110 kb. Exons are represented as blue boxes and the black line is the non-coding sequence. Human BAC clone RP11-626H4 was fragmented 47 kb downstream of the coding sequence by AscI/ClaI restriction enzyme digestion, and it was named human *PDGFRA* BAC-47kb.

4.2.3. Human *PDGFRA* BAC mice transgenic founders

After the pronuclear microinjection of the human BACs (Chapter 2.6), all the mice born were genotyped by amplifying the Human *PDGFRA* exon 1 by PCR (Chapter 2.2.14, data not shown). Seven founders of the Human *PDGFRA* BAC-110kb were found and labeled as #5,7,9,10,11,19,20 (Fig 4.3, A). In addition, 5 founders of the Human *PDGFRA* BAC-47 kb were named 1.3, 3.1, 3.4, 3.6, and 3.7 (Fig. 4.3, B).

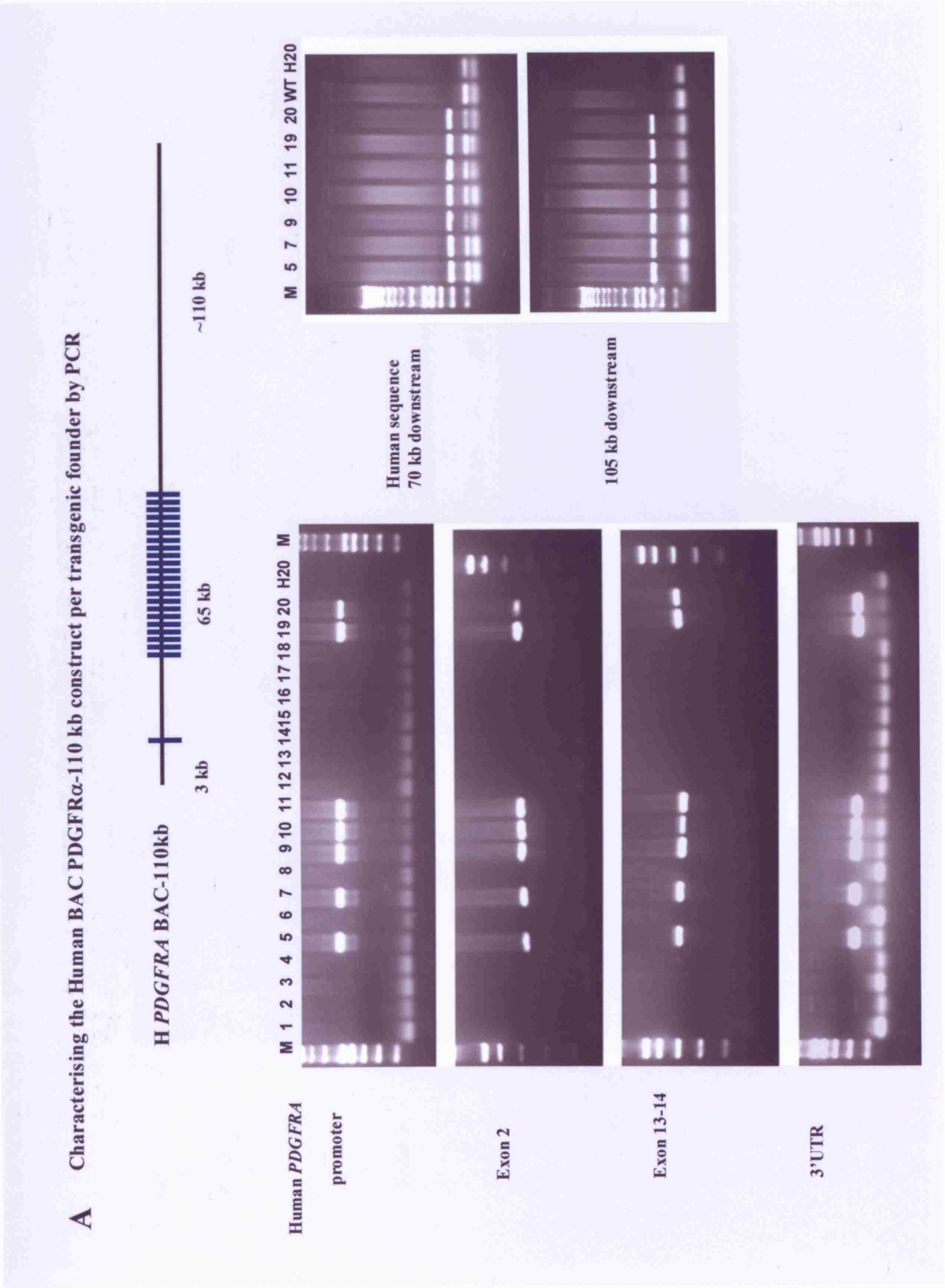
4.2.4. Checking the integrity of the BAC transgenes by PCR

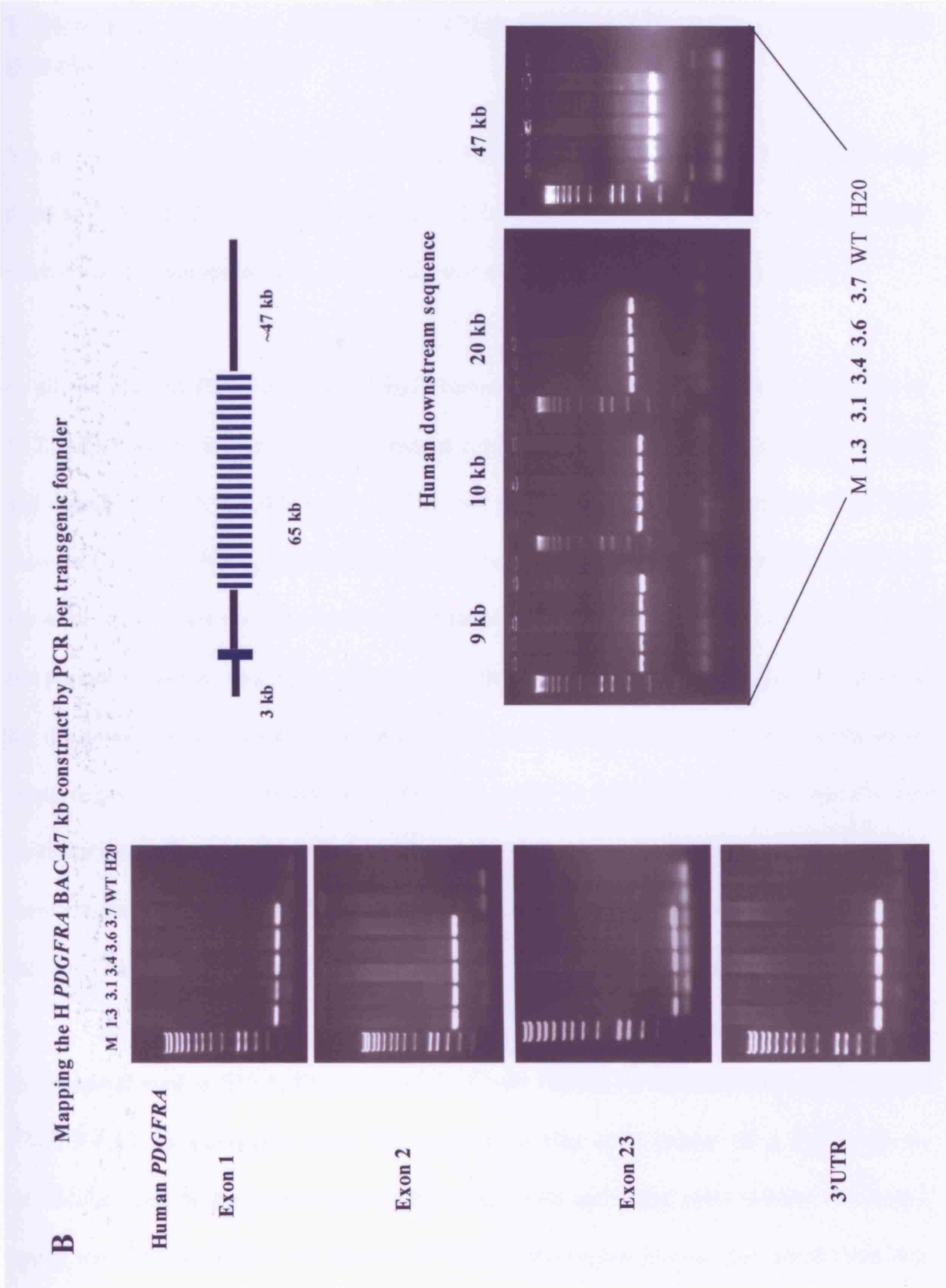
BAC DNA microinjected in the pronuclei of the fertilized egg could integrate complete or fragmented into the mouse genome, therefore the human *PDGFRA* gene and the downstream sequences were amplified from the genomic DNA of the transgenic pups. BAC transgenes were characterized, as described in Chapter 3 for the YAC transgenes, by PCR of the human promoter, human exon1, exon2, exon 13-14, exon 17, exon 23, 3'UTR; and 9, 10, 20, 47, 70 kb and 105 kb downstream sequence (Chapter 2.2.14, table 2.2 and 2.3).

I found that the Human BAC-110kb and Human BAC-47 kb constructs in all the founders seemed to have integrated complete in the mice genome (Fig. 4.3). All regions of the transgene were present in all the founders. All PCR reactions were specific for the human genomic sequences; they did not amplify the endogenous mouse *Pdgfra* gene.

These results show that the integration of the BACs occurred in more than one copy and did not fragment noticeably during integration. Higher copy number is presumably due to a higher concentration of BAC DNA isolated from the bacteria and easier purification method than the YACs. This then gives a huge advantage, in comparison with the fragility of the YACs which seem to fragment frequently.

Fig 4.3. Characterization by PCR of the Human *PDGFRA* BAC founders. A. Amplifications of the human promoter, exon 2, exon 13-14, 3'UTR, 70 kb and 105 kb downstream are shown in the seven founders of the Human *PDGFRA* BAC-110kb transgene (#5,7, 9,10,11,19,20). 7 from 20 pups contained the transgene, showing a 35 % efficiency of transgenic production. All the PCR amplifications were specific to human *PDGFRA* sequences. **B.** PCRs of the human exon 1, exon 2,exon 23, 3'UTR, 9 kb, 10 kb, 20 kb and 47 kb downstream are respectively shown for the five human *PDGFRA* BAC-47kb founders (1.3 3.1 3.4 3.6 3.7). Wild type genomic DNA (WT) and water only (H20) were used as negative controls, to detect any possible cross-amplification or contamination.


A Characterising the Human BAC PDGFRα-110 kb construct per transgenic founder by PCR
H PDGFRA BAC-110kb
3 kb
65 kb
~110 kb
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 H20 M
M 5 7 9 10 11 19 20 WT H20
Human PDGFRA promoter
Exon 2
Exon 13-14
3'UTR
Human sequence 70 kb downstream
105 kb downstream


B
Mapping the H PDGFRA BAC-47 kb construct by PCR per transgenic founder
M 1.3 3.1 3.4 3.6 3.7 WT H20
Human PDGFRA
Exon 1
Exon 2
Exon 23
3'UTR
3 kb
65 kb
~47 kb
Human downstream sequence
9 kb
10 kb
20 kb
47 kb
M 1.3 3.1 3.4 3.6 3.7 WT H20

4.2.5. Expression pattern analysis of the human *PDGFRA* BAC-110kb and Human *PDGFRA*-47kb transgenic mice at different developmental stages

The analysis of the expression of the human BAC-110kb and Human *PDGFRA*-47kb was done at E12.5, E13-13.5, E15.5, P0 and P21 in all the founders. The study was mainly concentrated in the spinal cord, the neural crest cells and surrounding sclerotome.

In all the Human *PDGFRA*-110 kb and Human *PDGFRA*-47 kb transgenic founders at E12.5, E13 and E13.5, expression in neural crest cells and sclerotome was again evident and strong (Fig 4.4, 4.5,4.6 arrows). This was predictable as all the previous YAC and plasmid Human *PDGFRA* transgenic founders representing 6 different transgenes showed the same strong expression in such cells at those stages of development. As discussed in the previous chapter, this study indicates that the *PDGFRA* gene expression is important for the proper development of the neural crest cells, sclerotome, and their derivatives in many organs during development. *PDGFRA* could be helping these cells initially for proliferation, migration and differentiation at the right places in the embryo. In addition, these results show again that the vector -either plasmids, YACs or BACs- are not affecting the specificity of the Human *PDGFRA* expression or related *LacZ* reporter gene.

In the spinal cord at E12.5, E13 and E13.5 of both Human *PDGFRA*-110 kb and Human *PDGFRA*-47 kb transgenic mice, the expression was at a cluster of a few cells in ventricular zone at the ventral part of the spinal cord and many cells outside the dorsal spinal cord VZ. At E15.5, the cluster of cells in the ventral part of the spinal cord has moved outside the VZ, the strong labelled dorsal cells were now distributed through the

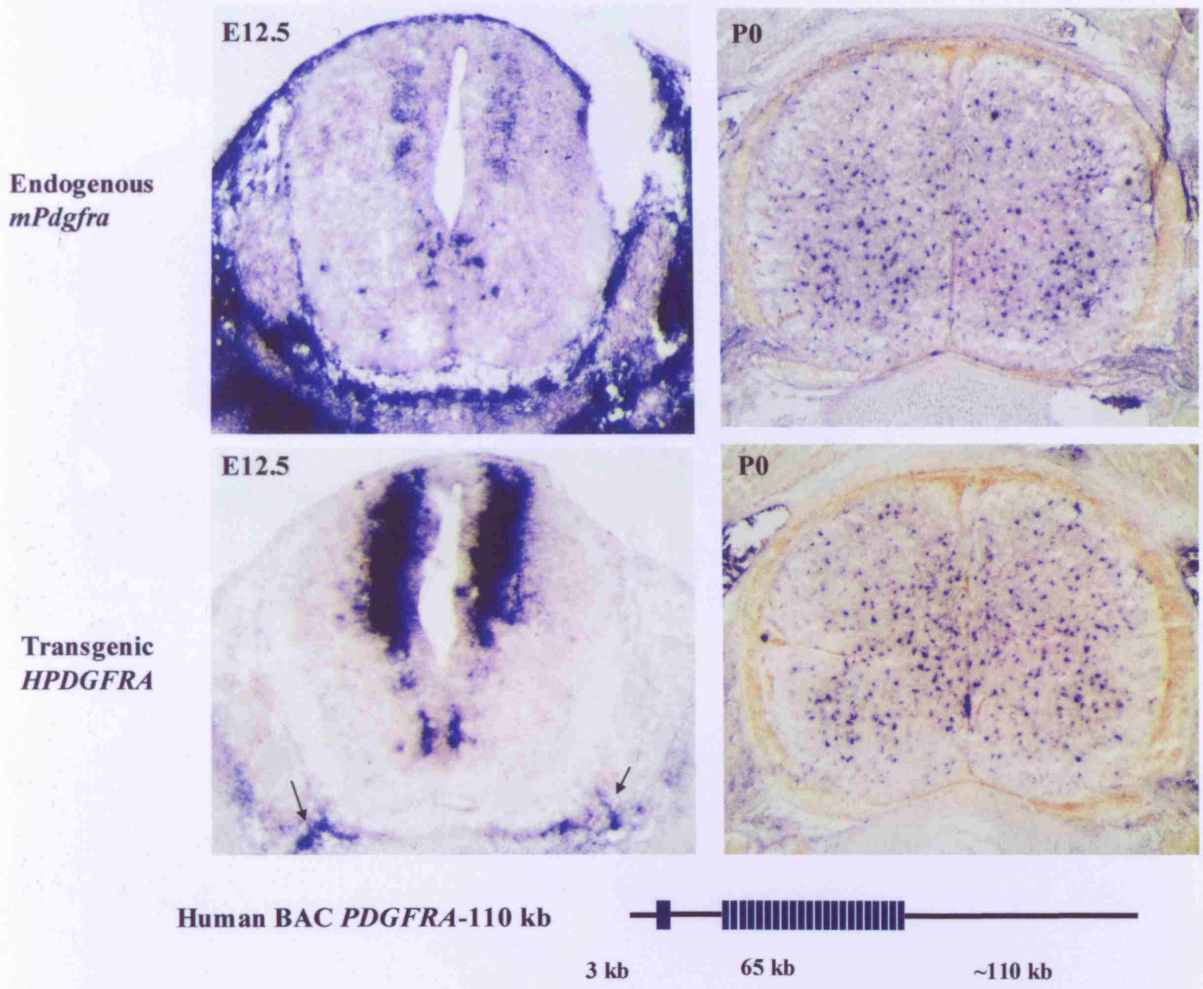


Fig 4.4. Comparative analysis of the Human *PDGFRA* BAC-110 kb transgene expression pattern and the endogenous *mPdgfra* gene in cervical spinal cord. Endogenous mouse *Pdgfra* expression pattern is shown at E12.5 and P0 for comparison. m*Pdgfra* is expressed in oligodendrocyte progenitors (OLPs), which are seen at the ventricular zone of the pMN ventral domain of the spinal cord at E12.5. At P0, OLPs are scattered throughout the postnatal spinal cord. In situ hybridization was done using a extracellular domain(ECD)-mouse *Pdgfra* probe. This probe binds properly to the m*Pdgfra* mRNA but also seems to bind to the transgenic human *PDGFRA* mRNA due to high homology in their ECD. This cross-hybridisation is very evident at E12.5, but not so much at P0. Below: cryostat transverse sections of E12.5 and P0 cervical spinal cord, in situ hybridization with the 3'UTR-Human *PDGFRA* probe does not cross-hybridise with the 3'UTR-m*Pdgfra* mRNA. At E12.5, there is strong expression in cells at the subventricular zone in most of the dorsal spinal cord, as well as a few cells in the ventricular zone at the ventral part. In postnatal spinal cords, human *PDGFRA*+ cells are widely distributed in all the P0 cervical spinal cord; there are also a few cells in the central canal strongly labelled. Their postnatal distribution seems to suggest that they could also be OLPs. Sclerotome progenitor cells and cranial neural crest cells are also strongly labelled surrounding all the spinal cord (arrows) at E12.5 and still at P0.

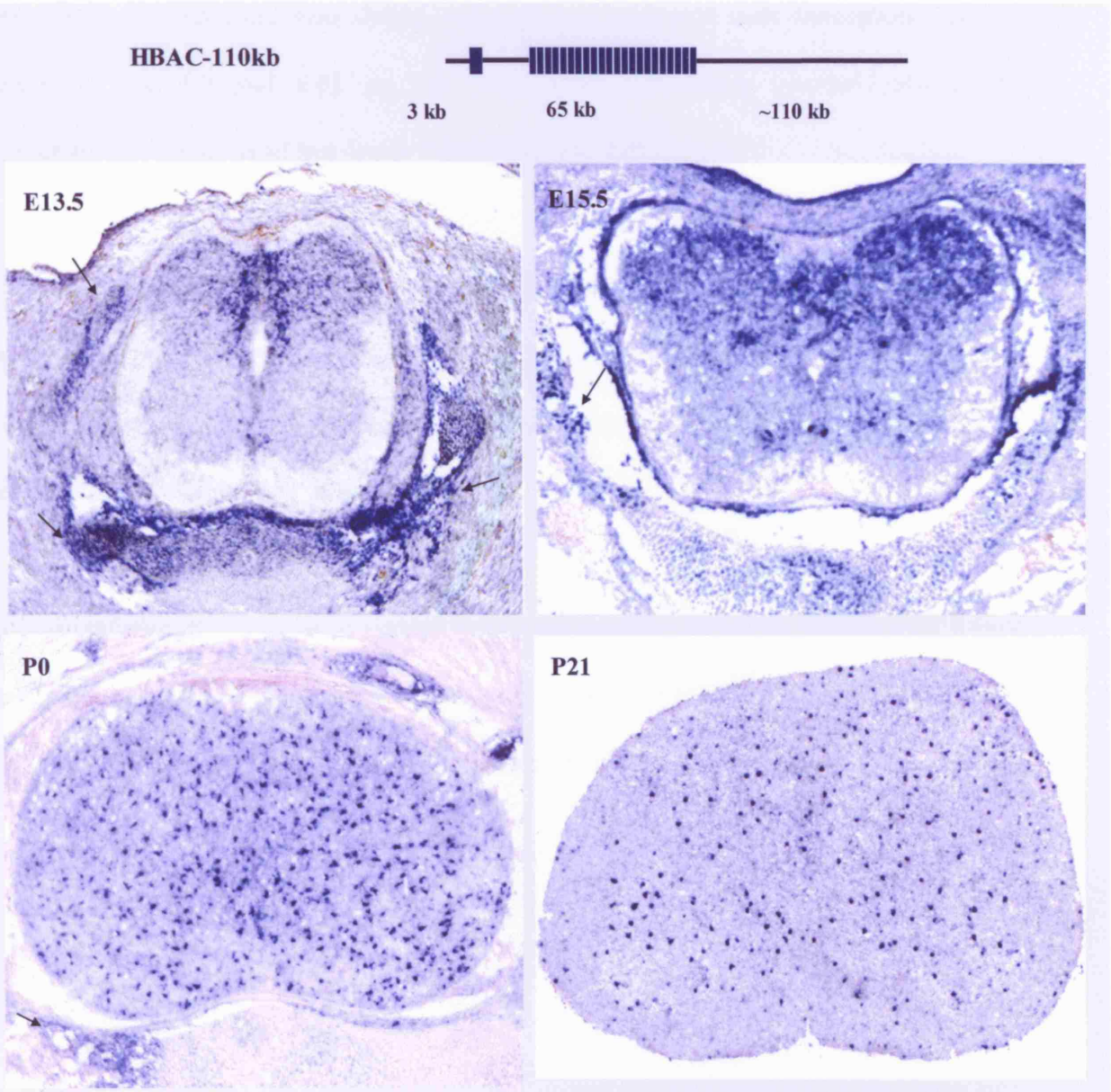


Fig 4.5. Analysis of the Human *PDGFRA* BAC-110 kb transgene expression pattern during cervical spinal cord development. In situ hybridization with the 3'UTR-Human *PDGFRA* probe in cryostat transverse sections of E13.5, E15.5, P0 and P21 cervical spinal cord. This probe does not cross-hybrise with the 3'UTR-m*Pdgfra* mRNA. Strong expression in cells at the subventricular zone in most of the dorsal spinal cord is evident at E12.5, as well as a few cells in the ventricular zone at the ventral part (Fig. 4.4). At E13.5, still a few cells in the ventral spinal cord keep expressing the transgene, however the dorsal cells are still strongly labelled, mainly distributed along the dorsal subventricular zone but many are now spread in the grey matter of the dorsal horns. This distribution continues at E15 but more cells are present, the ventral cells have also proliferated and migrate but much less than the dorsal positive cells. In postnatal spinal cords, now the human *PDGFRA*$^+$ cells are widely distributed in all the cervical spinal cord, at P0 there are also a few cells in the central canal strongly labelled. This postnatal distribution seems to suggest that they could be OLPs. At P21, there are still many human *PDGFRA*$^+$ cells but seem less in comparison the amount detected in the P0 spinal cord. Sclerotome progenitor cells and cranial neural crest cells are also strongly labelled surrounding all the spinal cord (arrows) at E12, E13.5, E15 and P0.

grey matter. At P0 there was strong expression in scattered cells throughout the spinal cord (Fig. 4.4, 4.5 and 4.6). At P21 there were still strong labelled cells scattered throughout the spinal cord but fewer than at P0 (Fig.4.5).

Different founders showed different intensities of expression but the cell type specificity appeared to be the same. Founder 3.1 of the Human *PDGFRA* BAC-47kb transgene showed no expression in the spinal cord and strong expression in the DRGs, but this pattern was peculiar to this one founder.

These expression patterns were similar to that obtained by Sun et al. (2000) with Y449C2. No significant differences among the three transgenes, Y449C2, Human *PDGFRA* BAC-110kb and Human *PDGFRA* BAC-47kb, were observed in the spinal cord.

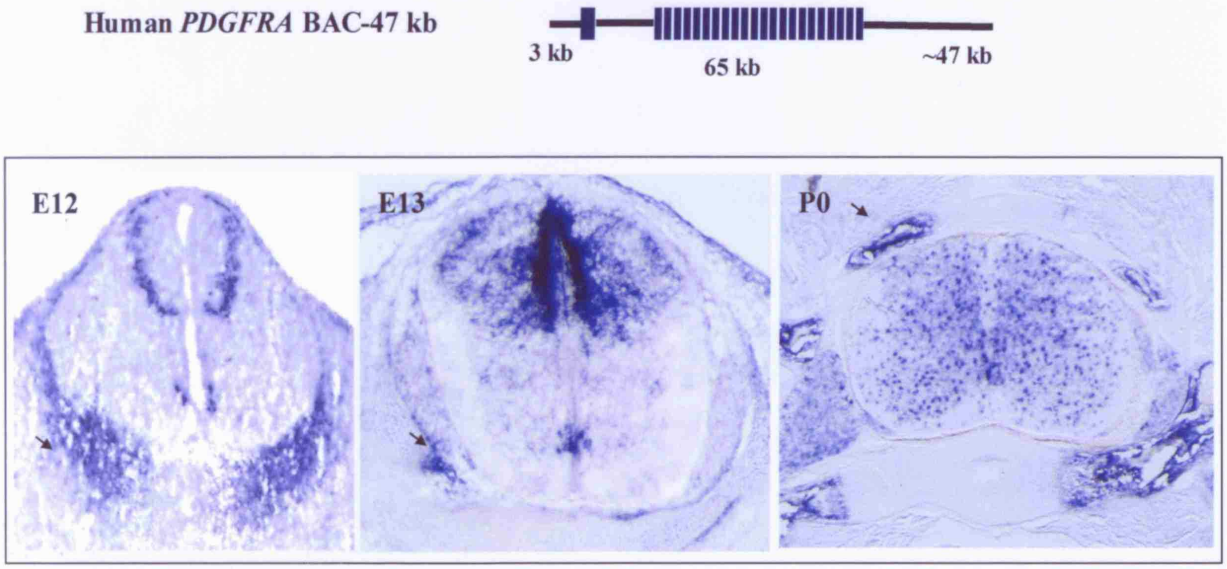


Fig 4.6. Expression pattern of the Human *PDGFRA* BAC-47 kb transgene in cervical spinal cord. In situ hybridization with the 3'UTR-Human *PDGFRA* probe in cryostat transverse sections of E12, E13 and P0 cervical spinal cords. At E12, there is expression in cells at the subventricular zone of the dorsal spinal cord, and in a few cells in the ventricular zone at the ventral part. At E13, the human *PDGFRA*$^+$ cells in the dorsal spinal cord are now closer to the ventricular zone, and there are also many distributed in the grey matter of the dorsal horns. A few cells also express in the ventral spinal cord. At P0, human *PDGFRA*$^+$ cells are widely distributed in all the cervical spinal cord, as well as some strongly labelled cells at the central canal. This distribution suggests that they are postnatal OLPs.

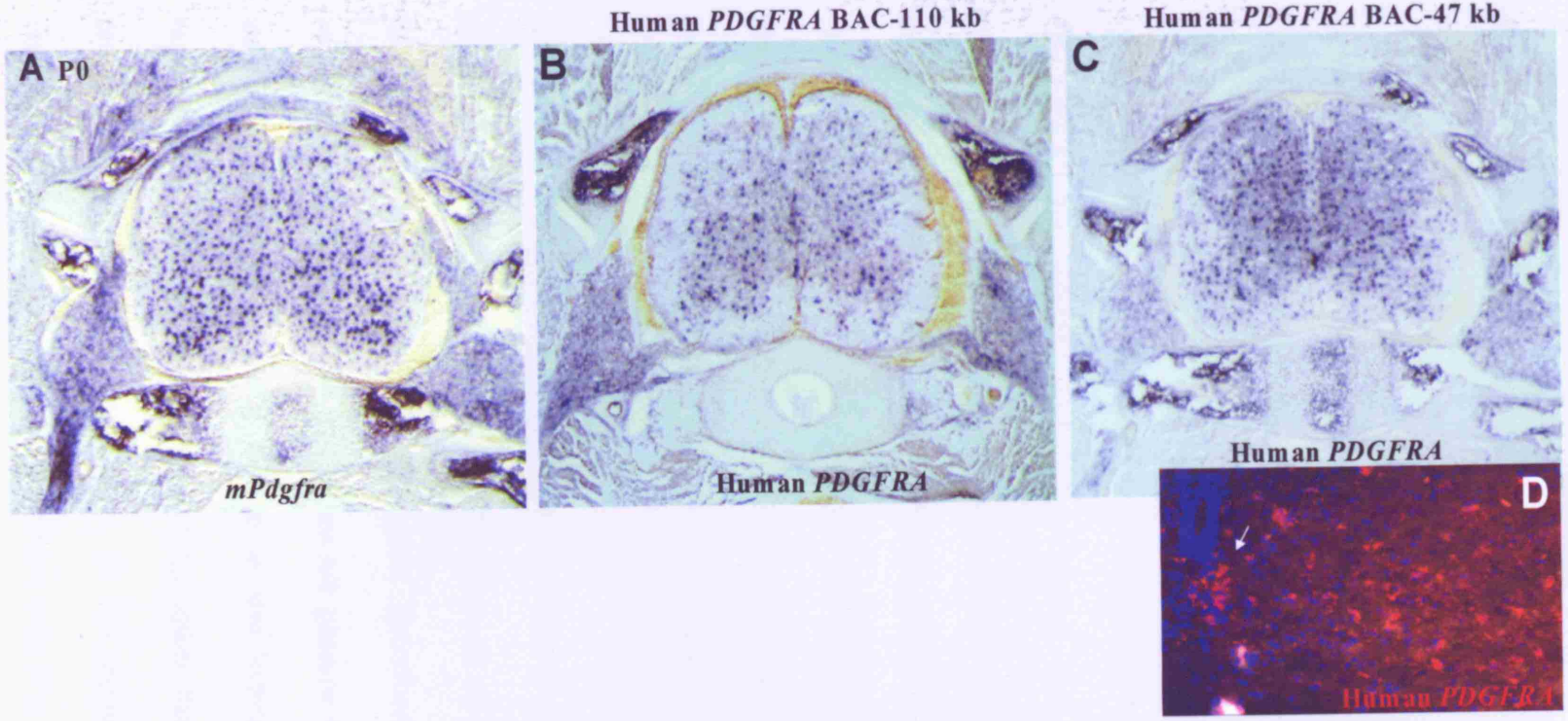


Fig 4.7. Comparison of the endogenous m*Pdgfra*, human *PDGFRA* BAC-110 kb and human *PDGFRA* BAC-47 kb transgene expression at P0 cervical spinal cords. A. Endogenous mouse *Pdgfra* is expressed in OLPs, they are scattered distributed in the entire spinal cord. B and C. The human *PDGFRA* BAC transgenes also showed a scatter distribution of human *PDGFRA*$^{+}$cells. In particular, there is strong human *PDGFRA* expression in a cluster of cells below the central canal in both transgenic lines. D. Close up to these cells (arrow) just below the central canal of a P0 human *PDGFRA*-47kb cervical spinal cord, fluorescent in situ using the same 3'UTR-human *PDGFRA* probe used in the other two human *PDGFRA* in situs. On the contrary, the endogenous m*Pdgfra* is not seen in those cells. A and C are consecutive sections. The Y449C2 transgene was similarly expressed in cells showing a scattered distribution and even in the cluster of cells below the central canal (Sun et al., 2000), to compare see fig. 4.10.

4.2.6. Expression of the Human *PDGFRA* transgene in OLPs; double in situ hybridization studies.

In order to prove the identity of the scattered cells expressing the Y449C2, Human *PDGFRA* BAC 110kb and Human *PDGFRA* BAC 47 kb, double in situ hybridization was carried out in transverse sections of E13, P0 and P4 spinal cords. At E13, double in situ hybridization for the oligodendrocyte transcription factor Olig2, and the Human *PDGFRA*-3'UTR mRNA probe was carried out to check the identity of the transgene-expressing cells at the ventral part of the spinal cord. The *Olig2* positive cells do not co-localize with the Human *PDGFRA* positive cells (Fig. 4.8), in fact they seem to fall exactly in the gap of expression between the ventral and the dorsal cells expressing the Human *PDGFRA* transgenes. This result showed clearly that OLPs at the pMN domain are not expressing any of the transgenes including the Y449C2.

However, at P0 another double in situ hybridization now between the mouse *Pdgfra* mRNA and the Human *PDGFRA* mRNA showed co-localization in all the cells, only a few appear to express one or the other gene. Therefore, it seems that at P0 many of these scattered cells expressing the transgenes are genuine OLPs (Fig.4.9). In addition, *Olig2* and Human *PDGFRA* mRNA double in situ hybridization shows co-localization of expression in most of the cells, proving again the identity of the oligodendrocyte progenitors expressing the transgenes. In this experiment, all the Human *PDGFRA* positive cells colocalized with *Olig2* positive cells (Fig.4.9), but some *Olig 2* positive cells did not coexpress the Human *PDGFRA* gene. Some of these cells could be differentiating oligodendrocytes, which rapidly down-regulate expression of *PDGFRA*.

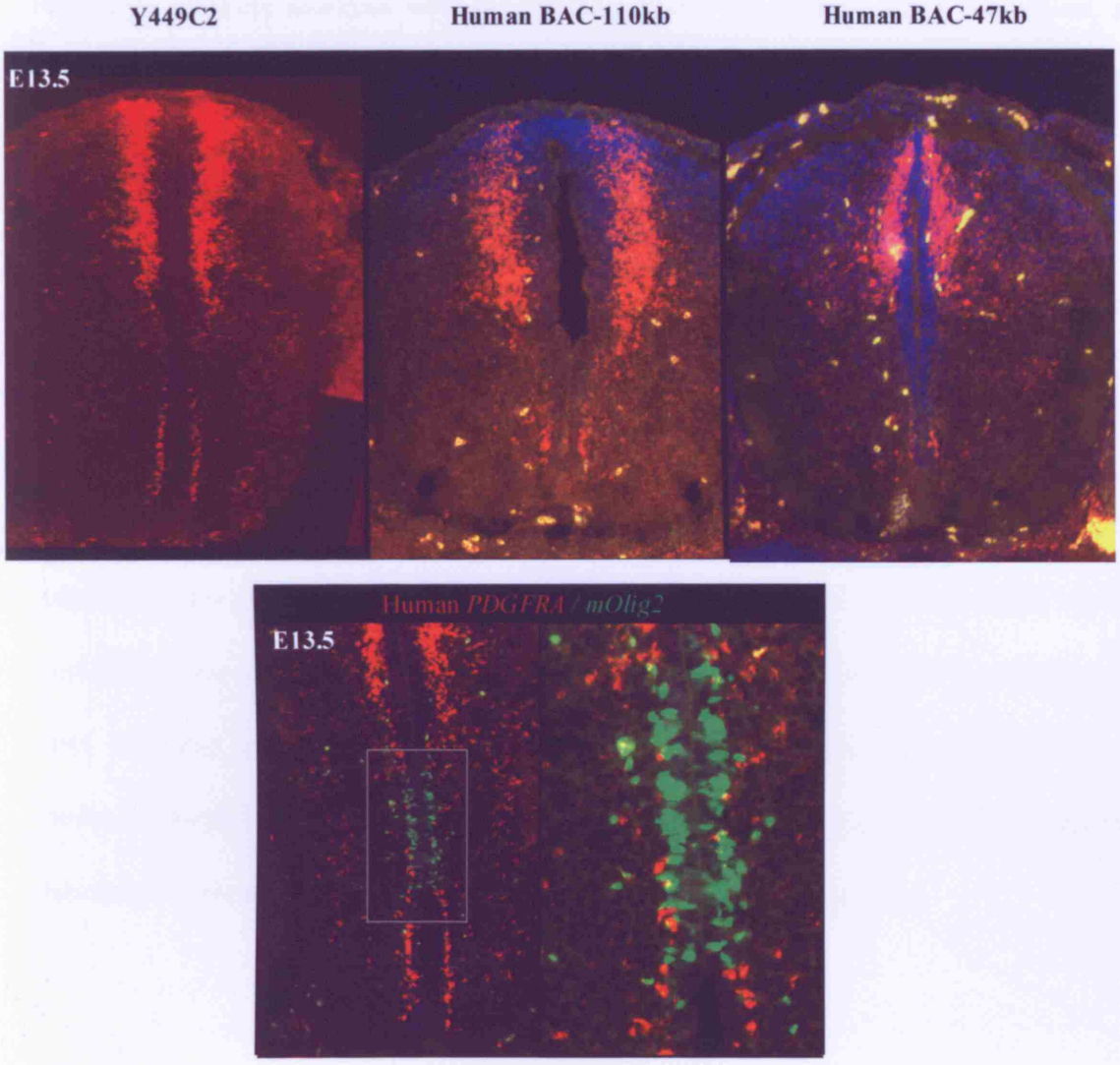


Fig 4.8. The Human BAC *PDGFRA* transgenes are expressed in postnatal OLPs, but not in early pMN ventral OLPs. Comparison of the expression pattern of Y449C2, human *PDGFRA* BAC-110 kb and human *PDGFRA* BAC-47 kb transgenes at E13.5 spinal cord. Fluorescent in situ hybridization with the 3'UTR-human *PDGFRA* probe in red and double fluorescent in situ hybridization with mouse *Olig2* in green. No colocalization of the human *PDGFRA* mRNA with the m*Olig2* mRNA was detected. It seems that in these three transgenic lines at E13.5, there is a gap of human *PDGFRA* expression exactly in the pMN domain (m*Olig2*$^{+}$ cells) of the ventral spinal cord. This is the domain where early OLPs appear and start proliferating.

Double in situ hydridization with the *NeuN* mRNA probe (Fig. 4.9) and immunohistochemistry with GFAP antibody show no colocalization of these neuron and astrocyte specific markers with the Human *PDGFRA* positive cells. Therefore, neurons and astrocytes do not express the Human *PDGFRA* transgenes in the P0 spinal cord.

Therefore, the oligodendrocyte progenitors seem to express the Y449C2, Human PDGFRA BAC-110 kb and Human PDGFRA BAC-47kb transgenes at P0, but not at E13.5 earlier in development.

It is possible that the activation of the Human *PDGFRA* transgenes occurs at downstream regions of the genomic sequence in order to get the specific OLP expression, but that this activation occurs postnatally but not during embryonic development. It could be possible that different regulatory elements or pathways of transcriptional activation are needed between embryonic and postnatal OLP development, and/or that these results represent the labeling of two different waves of OLP production in the spinal cord.

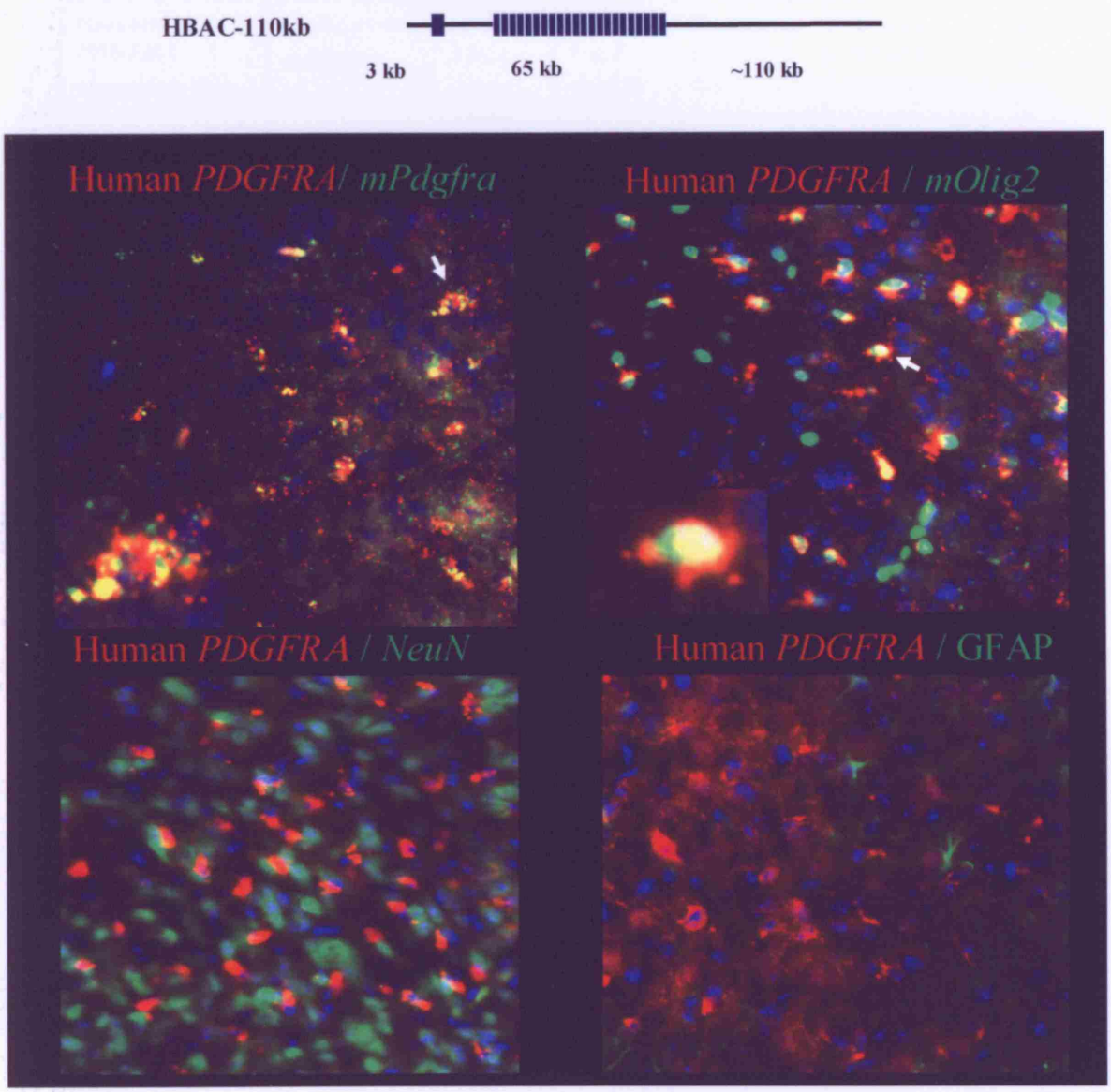


Fig 4.9. Expression of the Human *PDGFRA* BAC transgenes in oligodendrocyte precursors of P4 spinal cords. Human *PDGFRA* BAC transgene expression in oligodendrocyte precursors of P4 spinal cords. Double fluorescent in situ hybridization of the 3'UTR-human *PDGFRA* probe, versus mouse *Pdgfra,* mouse *Olig 2* and mouse *NeuN* probes, and GFAP antibody. Colocalization of the transgenic human *PDGFRA* mRNA with the endogenous mouse *Pdgfra* mRNA shows that such cells are postnatal OLPs as well as its colocalization with the OLP marker *mOlig2* (arrows) (close up to one cell in left bottom corner). Lack of colocalization with *NeuN* or GFAP shows that no expression of the human *PDGFRA* transgene was expressed in neurons or astrocytes respectively.

Human *PDGFRA* transgenic lines	**Endogenous *mPdgfra***	**2.2 kb-LacZ (Zhang, 1998)**	**Y29E11 LacZ exon14**	**Human *PDGFRA* BAC-47kb**	**Human *PDGFRA* BAC-110kb**	**Y449C2 (Sun, 2000)**
Neural crest cells	+	+	+	+	+	+
Sclerotome	+	+	+	+	+	+
1st and 2nd Branchial Arches	+	+	+	+	+	+
Mesenchyme: Lungs	+	+	-	+	+	+
Nostrils	+	+	-	**Nd**	**Nd**	+
Tongue	+	+	+	**Nd**	**Nd**	+
Heart valves	+	+	-	+	**Nd**	-
Kidney	+	+	-	**Nd**	+	+
Stomach	+	-	-	+	+	Nd
Intestine	+	-	+	+	+	+
Skin	+	+	-	+	**Nd**	+
Dermal hair papillae	+	+	-	+	**Nd**	+
Nervous system: Postnatal Oligodendrocyte precursors	+	-	-	+	+	+
Motor neurons	-	+	-	-	-	-
Cells in early dorsal spinal cord	-	+	+	+	+	+
DRGs	-	-	+	+	+	+
Trigeminal ganglia	-	+	+	**Nd**	**Nd**	Nd
Eye: Eye lids	+	+	+	**Nd**	**Nd**	-
Eye lens	+	+	-	**Nd**	+	+

Table 4.1. Expression pattern of human *PDGFRA* BAC transgenic lines. 2.2 Kb LacZ human *PDGFRA*, YAC29E11LacZexon14, human *PDGFRA* BAC-47kb and human

PDGFRA BAC 110-kb transgene expression patterns are summarized as well as the Y449C2 transgenic line and the endogenous mouse *Pdgfra* expression pattern included for comparison. Positive expression (+) or no expression (-) of the transgenes per cell type is determined after analysis of wholemount and/or cryostat sections with β-gal staining, or in situ hybridation with *LacZ*, 3'UTR-human *PDGFRA* or ECD-mouse *Pdgfra* probes. Not determined (Nd) expression means that it was not analysed in this particular stage or transgenic line.

4.3 Discussion

My human PDGFRα YAC transgenic mice lines (Chapter 3) confirm that the upstream sequence contains *cis*-specific regulatory elements (SREs) to drive the expression of this gene in cranial neural crest cells, sclerotome progenitors and mesenchymal cells. They also show that the long intron 1 and the sequence up to exon14 did not seem to contain other important SREs. Furthermore, none of those YAC transgenic lines showed expression in OLPs. Therefore, in this chapter I continue our search for SREs of the human *PDGFRA* gene, screening sequence regions containing the 3'end of the gene and downstream sequence. A human BAC clone, which contains the complete human *PDGFRA* gene, 3 kb upstream and 110 kb downstream sequence, was used as two constructs; one with 110 kb (Human *PDGFRA*-BAC-110kb) and the other with 47kb (Human *PDGFRA*-BAC-47kb) downstream sequence. After the analysis of seven transgenic lines of the Human *PDGFRA*-BAC-110kb, and five transgenic lines of the

Human *PDGFRA*-BAC-47kb at four developmental stages, 11 of them showed expression in postnatal OLPs in P0 cervical spinal cords. Both human *PDGFRA* BAC transgenes show expression in postnatal OLPs, as well as the already described expression in cranial neural crest cells, sclerotome progenitors and mesenchymal cells, but none seem to show expression in embryonic pMN OLPs.

In addition, the Human *PDGFRA*-BAC-110 kb and 47 kb transgenic constructs showed strong expression in many cells outside the VZ of the dorsal spinal cord and in a small cluster of cells in the ventral ventricular zone at E12.5 and E15.5 (Fig 4.4, 4.5, 4.6 and 4.8). At P0, both transgenes were expressed in OLPs scattered throughout the spinal cord, this expression continuing until p21 (Fig. 4.4, 4.5, 4.6 and 4.7). Their expression in postnatal OLPs was confirmed by the co-localization of mRNA from the transgenic human *PDGFRA* with the endogenous mouse *Pdgfra* and *Olig2* (Fig.4.9). The comparison of their expression pattern in transverse sections of P0 cervical spinal cord is shown in Fig. 4.7, together with the wild type expression of mouse *Pdgfra*. All show a scattered arrangement of cells in all the postnatal spinal cord, but only the transgenic P0 spinal cords show a small cluster of cells strongly labeled below the central canal (Fig 4.7, D). In general, the expression pattern of our Human *PDGFRA*-BAC transgenic lines during development were very similar to the Y449C2 transgenic line (Sun, 1999;Sun et al., 2000). Y449C2 was also expressed in cells below the central canal at P0 (Sun et al., 2000), however the reason of the expression of these three human *PDGFRA* transgenes in these cells is unknown.

A complete comparison of the P0 cervical spinal cord expression pattern of all our human *PDGFRA* transgenes, including the human YAC transgenic lines, is shown in Fig. 4.10, as well as the endogenous *Pdgfra* expression and the Y449C2 expression. Comparing the sequences of the YAC transgenes, adding the 23 kb long human *PDGFRA*-intron 1 and the gene sequence up to exon 14 (Fig. 4.10, II and III) does not contribute to the expression of the *PDGFRA* gene in OLPs. Nevertheless, the transgenes that contain the distal part of the gene and downstream sequence showed expression in OLPs at P0. The largest downstream sequence of around 310kb was included in the Y449C2 transgenic line. During this study, we fragmented the human BAC transgenes to test the reduction of the downstream sequence to 47 kb. The results show that the expression in postnatal OLPs continues even after deletion of 263 kb downstream sequence. This led me to consider that the *cis*-specific regulatory element of postnatal oligodendrocyte precursors (pOLPs-SRE) could lie in the sequence from intron 14-15 to 47 kb downstream of the coding sequence (Fig. 4.10). In total this sequence is around 63 kb long (Fig 4.11). At the begining of this project it was considered that this SRE for OLPs could lie anywhere in around 375 kb sequence, any sequence after human-*PDGFRA* exon1, therefore my study using YAC and BAC transgenes reduces this sequence around five times, eliminating 312kb sequences in total from the search for the SRE of postnatal OLPs.

In addition, as already discussed in Chapter 3, my results show that the SREs for cranial neural crest cells (CNCC), sclerotome progenitors (ScP) and mesenchymal cell derivatives (MsCD) are located in a small sequence of 3 kb upstream sequence (Fig. 4.11). All my transgenic founders from all the different human *PDGFRA* transgenes show expression in CNCC, ScP and MsCD, and all contain this minimum 3 kb upstream sequence (Fig. 4.10).

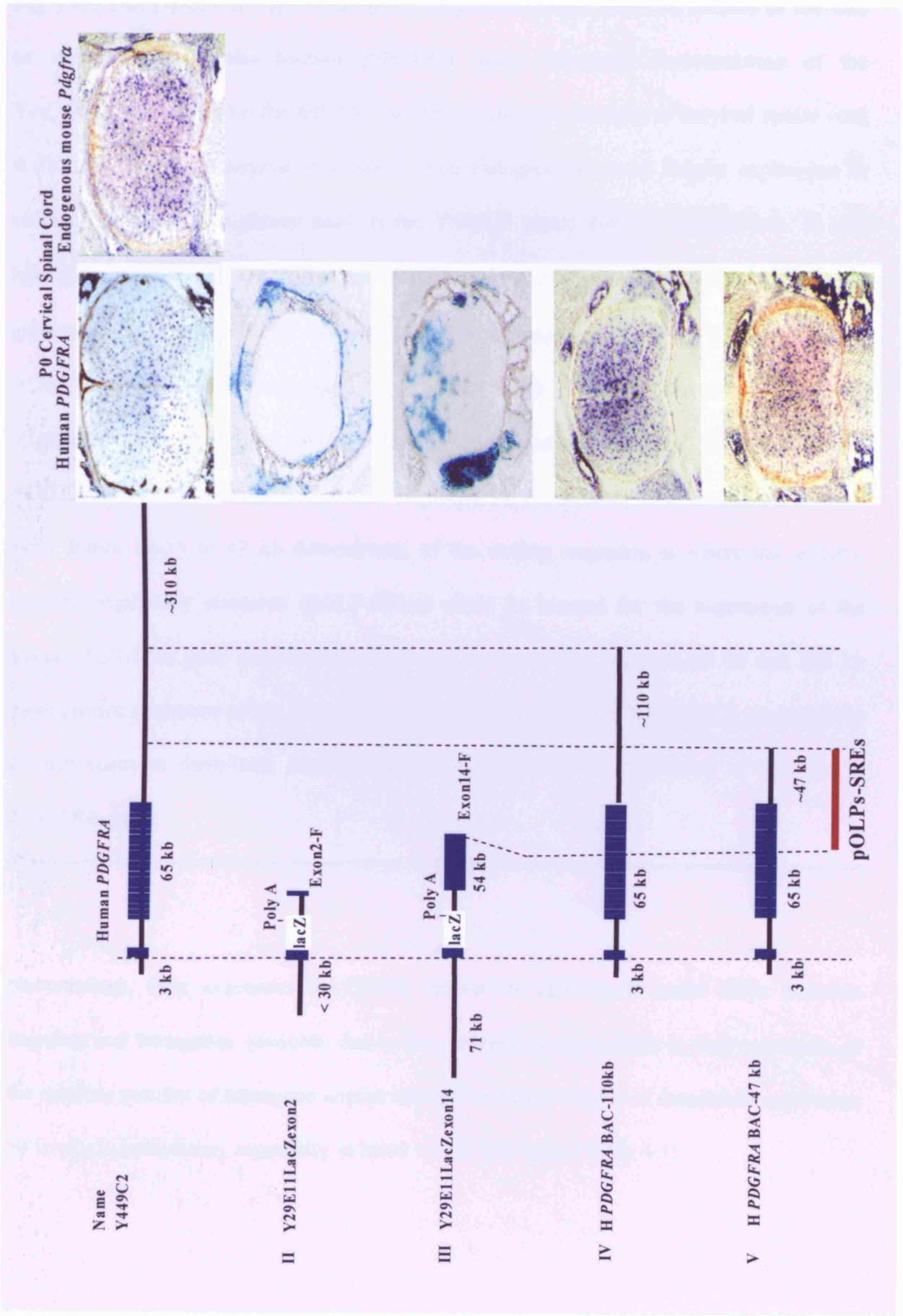

P0 Cervical Spinal Cord
Human PDGFRA
Endogenous mouse Pdgfra
Name
Y449C2
II Y29E11LacZexon2
III Y29E11LacZexon14
IV H PDGFRA BAC-110kb
V H PDGFRA BAC-47 kb
Human PDGFRA
~310 kb
65 kb
3 kb
Poly A
lacZ
Exon2-F
< 30 kb
71 kb
Exon14-F
54 kb
~110 kb
~47 kb
pOLPs-SREs

Fig 4.10. The postnatal oligodendrocyte progenitor-SREs could be located at the end or downstream of the human *PDGFRA* gene. Schematic representation of the YAC/BAC transgenes on the left. On the right, transverse sections of cervical spinal cord at P0 showing the transgene expression. The endogenous mouse *Pdgfra* expression in cervical spinal cord is shown next to the Y449C2 spinal cord for comparison. In situ hybridization with the 3'UTR-human *PDGFRA* probe in sections YAC449C2 (IV, V). β-galactosidase staining to P0 cervical spinal cord sections (II, III). Y449C2, human *PDGFRA* BAC-110kb and human *PDGFRA*-47kb were expressed in postnatal oligodendrocyte progenitors (pOLP). These three transgenes contained the entire human *PDGFRA* gene and a long downstream sequence. Approximately 63kb sequence, starting from intron 14-15 to 47 kb downstream of the coding sequence, is where the pOLPs-specific regulatory elements (pOLP-SREs) could be located for the expression of the human *PDGFRA* gene (marked as a red line). It seems that the extra 63 kb and 263 kb downstream sequence of the human *PDGFRA* BAC-110 kb and YAC449C2, respectively do not seem to contribute importantly to the regulation of expression of the human *PDGFRA* gene.

Nevertheless, their expression in CNCC derivatives and MsCD could differ between founders and transgenes, probably due to the presence of small SREs in their sequence, or the random number of transgene copies inserted and their degree of detectable expression by in situ hybridization, especially in heart valves and lungs (Table 4.1).

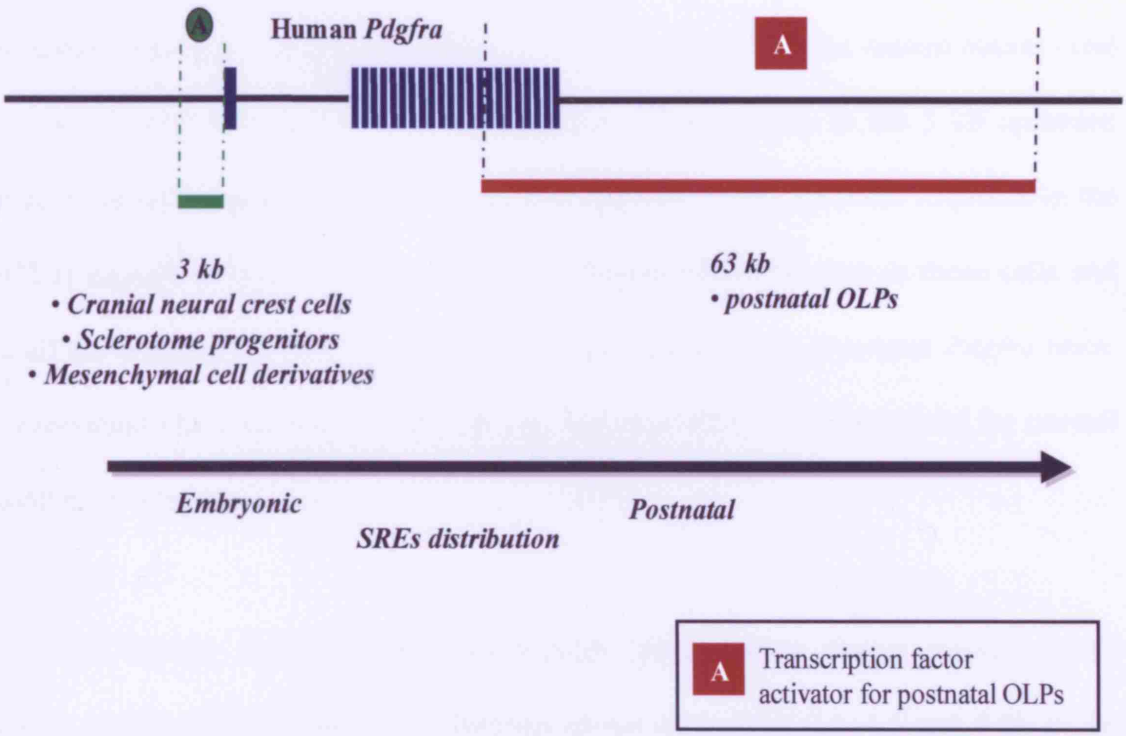


Fig 4. 11. Model suggesting the location of some *cis*-acting regulatory elements for the Human *PDGFRA* gene. As a result of the expression pattern comparison of ten human *PDGFRA* transgenic lines, it is now possible to suggest the non-coding region where specific transcriptional regulatory elements (SREs) of the human *PDGFRA* gene could be found for several progenitor cells. The human *PDGFRA* transcription was activated in cranial neural crest cell, sclerotome progenitors and mesenchymal derived cells in all the transgenic lines. It seems that only including 3 kb of upstream sequence is enough to activate the expression of *PDGFRA*. Therefore, important specific regulatory elements to activate the transcription of human *PDGFRA* gene in those cells must be included in that 3 kb sequence. Furthermore, this research has narrowed down the non-coding region where the SREs for postnatal OLPs could be found, to a 63 kb sequence from intron 14-15 to 47kb downstream sequence. Early OLPs-SREs could not be located, because their expression was not detected in any transgenic line, which could mean a location Mb far from the gene or a *trans*-regulatory control system. SREs for the expression of the human *PDGFRA* gene in mesenchymal derived cells of specific organs, like lungs, intestine, kidneys, were difficult to locate due to the complexity of the transcriptional regulation control of the gene, in some cases it showed that specific repressors could be present as well as activators; this requires to be studied further.

Sun et al. (2000) demonstrated that the Y449C2 transgenic line can rescue the knockout *Pdgfra* mice up to birth, rescuing especially their craniofacial and bone structure abnormalities (Sun et al., 2000). According to my data localizing the cranial neural crest cells, mesenchymal cells and sclerotome progenitor SREs mainly in the 3 kb upstream sequence (Chapter 3 and Fig. 4.11), it seems that only the 3.5 kb upstream sequence in the Y449C2 is enough to activate expression of the human *PDGFRA* gene in those cells and rescue all the craniofacial and bone structure abnormalities in the knockout *Pdgfra* mice. This observation identifies the 3.5 kb upstream region of *PDGFRA* as essential for normal mammalian development.

Both of our human *PDGFRA* BAC transgenics also showed strong ectopic dorsal expression in E12.5, E13.5 and E15.5 cervical spinal cords (Fig. 4.4, 4.5 and 4.6). In the Y449C2 transgenic mice, it was interpreted that this strong dorsal expression in E12.5 and E15 spinal cord could be due to the ectopic activation by *KIT* (Sun, 1999;Sun et al., 2000), due to the interfering effect of a spinal cord-specific enhancer element possibly located between 38kb and 146kb upstream of the *Kit* promoter (*PDGFRA/KIT* intergenic region is 359.7kb long, UCSC Genome Browser Mar. 2006 assembly). The presence of a KIT enhancer in this region is presumed because of the observation that in the W^{57} mouse mutant spinal cord *Kit* expression seems to be abolished (Kluppel et al., 1997;Berrozpe et al., 1999;Sun et al., 2000). Y449C2 transgenic line could possibly contain this enhancer as this YAC extended up to 87 kb from the *KIT* promoter (Sun et al., 2000). Assuming this, it could also be possible that Y29E11 contains this enhancer due to its 310 kb downstream sequence. However, the human *PDGFRA* BACs that we used could not contain this *KIT* enhancer due to their shorter 110 and 47kb downstream sequences. This shows that the

interpretation that the ectopic dorsal expression in the human *PDGFRA* transgene spinal cords is due to the activation by that *KIT* spinal cord-specific enhancer is no longer viable. Two alternative explanations for this ectopic activation of the human *PDGFRA* transgene in those dorsal cells of the spinal cord are: 1) It could be due to the mis-regulation of the sequence by loss of a specific repressor element or 2) the extra number of copies of the transgene could somehow cause loss of correct regulation, e.g. by titrating out a repressor protein.

Some information about the expression of the *PDGFRA* gene in heart valves and lungs was obtained, but not about its regulation because the expression in these organs was activated inconsistently between founders, transgenes and developmental stages. More repetition and specific analysis on these tissues is required.

Even though the postnatal OLPs express both human BAC transgenes, the expression in early pMN-derived OLPs (pMN-OLPs) was not so clear. It was easy to think that the cluster of cells in the ventral part of the transgenic spinal cords (Fig. 4.4, E12.5; Fig 4.5, E13.5, Fig. 4.6, E12 and E13) are pMN-OLPs, however with double fluorescent in situ hybridisation I now have shown that the OLP marker *Olig 2* did not co-localise with the transgenic human *PDGFRA*. We also now know that Y449C2 is not expressed in early OLPs in pMN as originally thought (Sun et al., 2000). This also shows that none of the previous human *PDGFRA* transgenic lines and none of our new transgenes induced their expression in pMN-OLPs. It is surprising that by inserting the entire human *PDGFRA* gene, 310 kb downstream and around 60 kb upstream sequence, 447 kb genomic sequence in total considering all the YAC/BAC constructs in this project, we could not locate the

pMN-OLP SREs. Nevertheless, regulatory elements have been found upstream of the *KIT* gene, located between -23kb and -154kb from the *KIT* promoter (Berrozpe et al., 1999;Berrozpe et al., 2006). Therefore, it could still be possible that the activation of for the 65 kb long *PDGFRA* gene transcription in the pMN-OLPs could be directed by SREs located even further upstream than 60 kb. There are two genes upstream of *PDGFRA* gene, *GSH2* and *CHIC2,* but the intergenic sequence between human *GSH2* and *PDGFRA* exon1 is 127. 324 kb long, and between *CHIC2* and *PDGFRA* exon1 is 164.663 kb long, according to the UCSC Genome browser (Mar. 2006 assembly). Consequently, there is only a ~67 kb extra upstream sequence possible to search for pMN-OLP SREs.

Furthermore, the discovery of a global control region (GCR) that regulates the expression of the *HoxD* gene cluster (Spitz et al., 2001;Spitz et al., 2003;Spitz and Duboule, 2005), raises the possibility that GCRs could exist in other mammalian gene clusters, including the *PDGFRA-KIT-KDR* cluster. The *PDGFRA-KIT-KDR* cluster is located on human chromosome 4 q12 (Gronwald et al., 1990;Kawagishi et al., 1995), from 54790203 to 55686510 (UCSC Genome Browser) and is 896.307 kb long. On mouse chromosome 5, *PDGFRA-KIT* and *FLIK1* are acting as a functional cluster that helps to define 5 subsets of mesodermal cells during early development (Kataoka et al., 1997).

If the human *PDGFRA* transcription in early pMN-OLPs depends on *cis*-regulatory elements, such as a far more upstream sequence, a GCR sequence, or shared SREs of neighbouring genes that were not included in any of my *PDGFRA* constructs, then more experimental sequence analysis needs to be done to find the pMN-OLP SREs and/or replicate the regulatory conditions for their activation.

It is also possible that the regulation of the expression of this gene in pMN-OLPs depends on other control systems for its transcription, such as the three dimensional organization of the gene by loop formation or distance to the nuclear matrix (Bulger and Groudine, 1999). These regulatory mechanisms cannot be completely replicated or guaranteed in our transgenic system, because in some situations we are inserting only fragments of the original gene sequence, but mainly because we have no control over the site of integration of our long human *PDGFRA* constructs in the mouse genomic DNA. It would require a challenging experimental model to replicate such conditions, and study the importance of three-dimensional regulation.

During the Y449C2 rescue of the knockout *Pdgfra* mice (Sun et al., 2000), it now seems that the early pMN-OLPs were not rescued but the postnatal or late OLPs were, so the observed *in vitro* and *in vivo* recovery of OLPs from the rescued spinal cords showed the expression of the human *PDGFRA* in postnatal OLPs but not the complete rescue of *Pdgfra* expression in all OLPs. However, the postnatal OLP wave could be enough to recover and replace the final OL population in the spinal cord. It will not be possible to probe the complete recovery of OLs in the rescued pups because they die immediately after birth due to respiratory failure.

The data raise the possibility that postnatal OLPs labelled in the human *PDGFRA* transgenic mice are not all derived from the ventrally-pMN derived OLPs, but possibly from a dorsally derived or second wave of OLP production. It would still be very interesting to analyse by *in situ* hybridization, human *PDGFRA* BAC-47kb embryos at

E16, E17 and E18.5 to check exactly when and where transgene starts expressing in OLPs, and whether the postnatal OLPs originate from the ventral or dorsal part of the spinal cord. The results of this in situ hybridisation analysis could show if these human *PDGFRA* BAC transgenic mice could be used for the specific study of dorsal, postnatal, or a second developmental wave of OLPs, and add to the discusion about the site of origin of OLPs.

In conclusion, my human *PDGFRA* transgenic analysis suggests that SREs are located sequentially in the non-coding genomic sequences, depending on their developmental time of activation. Early developmental expression-SREs seem located in the upstream sequence, such as CNCC-SREs, and late developmental expression-SREs, such as pOLP-SREs, seem located either in the distal part of the gene or in the downstream sequence (Fig. 4.11). Therefore with these results, I suggest a model of 'temporal collinearity' of activation of the *PDGFRA*-SREs during development (Fig.4.11, black arrow) and that the transcriptional regulatory controls for this gene are possibly divided into embryonic and postnatal regulatory systems.

Similar models have been recently suggested for the *HoxD* gene cluster (Spitz et al., 2001;Spitz et al., 2003;Spitz and Duboule, 2005) and the 93DE cluster in *Drosophila* (Jagla et al., 2001), considering that this "temporal colinearity" could be a mechanism to coordinate the expression of genes in time. The *PDGFRA* gene has three possible promoters, 23 exons, is 65 kb long and produces four transcripts, it is possible that mammalian genes with these caracteristics is organized for its transcriptional expression similarly to clusters of small genes as the Hox clusters.

Chapter 5.

In silico analysis of the *pdgfra* gene during vertebrate evolution

5.1 Introduction

5.1.1. Gene regulatory systems during evolution

All biological processes are driven by the information stored in our DNA. Part of the information for life processes is encoded in genes, which once translated produce proteins that are essential for our cells (Fig. 1.3). These genes are formed, conserved, arranged, changed or deleted over billions of years of evolution. These genes or coding regions of our DNA have been extensively studied. Many genes of different species keep highly conserved, making possible to make human genes to function in bacteria, yeast or a fly cell, so it seems that subtle changes in its DNA sequence are the main contribution to evolution and are favour during adaptation processes. However, the non-coding regions, sequences of DNA between or inside genes that are not used to codify a protein, have not been studied at the same rate because they were first considered unimportant just as DNA sequences that keep the genes separated or are used for packing as chromatin. Nevertheless, these non-coding sequences contain the binding sites of proteins that regulate gene expression during all phases of an organism's life. It is very probable that these regulatory sequences in the genome are also subject to evolutionary pressures, being conserved or mutated in a similar way as protein-coding sequences.

Genome sequences, protein sequences, in situ hybridisation gene expression patterns and many other databases are now available freely on the internet. Sequence genome projects for *C. elegans*, yeast, fungi, bacteria, zebrafish, Xenopus, chicken, mouse, pig, dog, chimpanzee and human are finished or close to being finished. The application of mathematical and computational modelling is helping to work faster with the enormous amount of biological information already accumulated, and to predict the transcriptional regulatory networks of important eucaryotic genes by bioinformatical analysis (Gibson and Honeycutt, 2002;Pritsker et al., 2004). The comparison of coding gene sequences between mouse and human is now available in GenBank and comparisons between coding sequences of different species could be simple, using the long genome multi-alignment tools already available. These tools allow us to compare not only the coding sequences of genes but also their flanking sequences, potentially allowing us to identify conserved regulatory sequences in the non-coding sequence of several species, providing a prediction for the location of *cis*-acting regulatory elements such as those I have been searching for experimentally.

Already several studies have compared DNA sequences among evolutionarily distant genomes with bioinformatic tools to identify conserved functional regions in non-coding DNA. Santini et al. (2003), for example, compared the *HoxA* clusters of tilapia, pufferfish, striped bass, zebrafish, horn shark, human and mouse which span approximately 500 million years of evolution. Many identified putative regulatory elements that are known to regulate the expression of *Hox* genes as well as putative new elements were uncovered. The majority of the newly identified putative regulatory

elements contain short fragments that are strongly conserved and contain known binding sites for regulatory proteins. They also noticed that the regulatory intergenic regions located between the genes that are expressed most anterior in the embryo are longer and apparently more evolutionarily conserved that those at the other end of *Hox* clusters. This suggested that the conserved elements are involved in different gene regulatory networks and supports the duplication-deletion-complementation model of functional divergence of duplicated genes (Santini et al., 2003).

In this chapter, I analyse the predicted structure of the *pdgfra* gene of zebrafish, Xenopus, fugu, dog, chicken, mouse, chimpanzee and human, and describe the homologies in coding and non-coding sequences through multiple alignments, showing some interesting conserved sequences (CS) among these vertebrate. The use of these data complemented with a list of transcription factor binding sites (TFBS) found in those conserved sequences provide information that can be used to propose the first *in silico* map of transcriptional regulatory complexes of the *pdgfra* gene.

5.2 Results

Search for conserved sequences (CS) and transcription factor binding sites (TFBS) in the non-coding sequences of the *pdgfra* genes of different vertebrates.

5.2.1. Analysis of CS in non-coding mouse and human upstream and downstream *pdgfra* sequence using BLAST multi-alignment, TRANSFAC® database and MATCH ™ tool.

Transcription in eukaryotic cells is regulated in a complex way through a system of interactions of transcription factors (TF), whose activation or repression activities are determined by their DNA binding sites (TFBS) on a target gene. Each cell type at a developmental stage has its own gene expression profile that is in part defined by the presence of specific combinations of TFs in the cell (Matys et al., 2003).

In order to identify the regulatory signals in new genome sequences, databases have been established, maintained and updated, focussing on different aspects of transcriptional control mechanisms. A database developed in the 1980s is TRANSFAC®, which is a database of transcription factors and their binding sites in genomic DNA. TRANSFAC® can be freely accessed on the internet at http://www.gene-regulation.com. The last updates and new tools have led to a better understanding of tissue-specific expression of genes by sub-dividing the TF databases in tissues or organs (Wingender et al., 2001;Matys et al., 2003). The MATCH ™ tool, for example, was developed to search any genomic

DNA sequence for potential TFBS (http://www.gene-regulation.com/cgi-bin/pub/programs/match).

Simultaneously with the human *PDGFRA* transgenic analysis, I started a bioinformatic *pdgfra in silico* analysis, searching for conserved sequences (CS) in non-coding mouse and human sequence using first the BLAST multi-alignment tool (http://www.ncbi.nlm.nih.gov/BLAST/), and then the TRANSFAC® database with the MATCH ™ tool. During this analysis, I aligned 71 kb upstream sequence, the entire intron 1 and 98 kb downstream sequence of both human and mouse *Pdgfra* sequences. The alignments and TFBS search were done separately for upstream, intron 1 and downstream sequences. I identified TFBS in conserved sequences by hand, comparing the results of BLAST multi-alignment and the TFBS sites found with MATCH™ on each sequence. In total, 73 CS of 99% identity between human and mouse and with different lengths, from 10s to 100s of bps, were found in the 71 kb upstream region; 18 CS were found at intron 1, and 48 CS in the 98 kb downstream sequence. TFBS were found with MATCH, using the available TFBS database for vertebrates, finding hundreds of hits in the human and mouse sequences. A schematic representation of the TFBS found in CS of human and mouse in each region is shown in Fig. 5.1. Surprisingly, there were no TFBS found in CS of the long intron 1. However, these results depend entirely on the type and number of TFs listed in the TRANSFAC database so is limited in value.

Of the TFBS identified in this preliminary screen, FoxJ2 and Pax4 have been reported to be expressed in the CNS – and only Pax4 in the ventral spinal cord. On the other hand,

FoxD3 could be important for *pdgfra* expression in neural crest cells during early development.

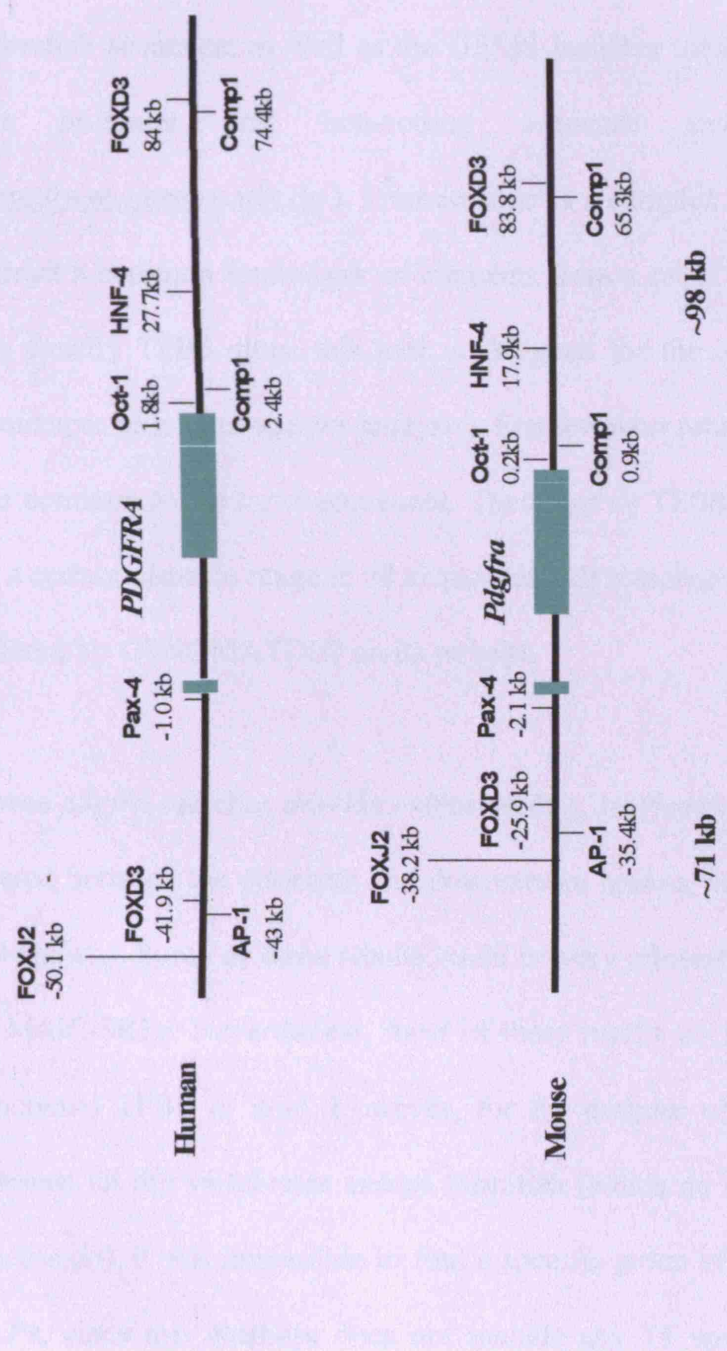


Fig. 5.1. Transcription factor binding sites (TFBS) in human and mouse conserved sequences (CSs) surrounding the *pdgfra* gene. 71 kb upstream sequence, the entire intron 1 and 98 kb downstream sequence were aligned with BLAST (99% identity between sequences). For each human or mouse sequence, TFBS were found with MATCH using the TRANSFAC database. Only the TFBS that were found in CSs are listed in this schematic representation. Its localization either in human or mouse sequence is measured considering for upstream the beginning of exon1 as 0, and for downstream the end of exon23 as 0. There are not transcription factor binding sites predicted at the conserved sequences of intron 1. All these TFs could be used for the specific transcriptional regulation of *pdgfra*, and especially *FoxD3, Pax4, Oct-1,FoxJ2* and *HNF-4* during mammalian development. *Pdgfra* exons are represented as green boxes and the non-coding sequence as a black line.

5.2.2. Di-Align, MatInspector and Frameworker from GENOMATIX ©.

Bioinformatic analysis of 100 kb upstream and downstream regions of *PDGFRA* were also done using, the Clustal X and DiAlingTF multi-alignment of human, mouse, chicken and zebrafish sequence; as well as the GEMS launcher tools MatInspector and Frameworker for promoter and non-coding sequence analysis from GENOMATIX© (http://www.genomatix.de). Frameworker is a complex software tool that allows users to extract a common framework of elements from a set of DNA sequences. These elements are usually TFBS since this tool is designed for the comparative analysis of promoter sequences (e.g. inter-species analysis). Frameworker returns the most complex models that are common to the input sequences. These are all TFBS that occur in the same order and in a certain distance range in all sequences. All searches were done with the free trial tools offered by GENOMATIX© on its website.

These *pdgfra* searches provide extensive data, hundreds of hits for TFBS and even TFBS shared between the upstream and downstream human, mouse and chicken sequence (data not shown). Some of these results could be very relevant to find CNCC-SREs, ScP-SREs or MsdC-SREs. Nevertheless, most of these results could be false positive TFBS or not functional TFBS *in vivo*. However, for the purpose of finding a SRE of OLP shared between all the vertebrates except zebrafish (which do not express *pdgfra* in OLPs, see discussion), it was impossible to find a specific group of TFBS that could be relevant for OLPs, since this database does not include any TF specific during OLP development. Only the E2A proteins, and specifically E12 could be consider as an interesting hit (data not shown), since it is been described that Olig2 uses E12 and E47 to bind to the DNA

sequences (Samanta and Kessler, 2004), and there is still the possibility that it could be used to regulate *pdgfra* transcription.

5.2.3. Multiple alignments from the UCSC Genome browser

Only three to four years ago, some browsers started offering conservation comparisons between long non-coding genome sequences; however they were not reliable due to the incomplete genomic sequences, alignments, gene information and mathematical modeling. However, one of the most complete was the UCSC Genome browser. In the UCSC Genome browser human May 2004 assembly (in http://genome.ucsc.edu/cgi-bin/hgGateway), three short CS between the *PDGFRA* gene and the *KIT* gene were found by comparing the genomic sequences of different species. One of them, CS#1 is 213 bp long and is located ~24.8 kb downstream of the human *PDGFRA* 3'end, so it lies inside the 47 kb downstream sequence identified by our transgenic analysis (Fig. 5.2). In order to investigate whether CS#1 is a SRE of the human *PDGFRA* gene and if it could be specific for the expression in postnatal OLPs, I followed two experimental approaches, one deleting this CS#1 from the human *PDGFRA*-BAC 110 kb by homologous recombination (named CE#1^{-} HBAC-110 kb), and the other by cloning it into the 2.2kb LacZ-human *PDGFRA* plasmid (named CE#1^{+}H2120-βgal) (Zhang et al., 1998) (Fig. 3.13), downstream to the LacZ sequence. Both CS#1 constructs were used to produce transgenic mice by pronuclear microinjection, 5 founders for each construct were analysed. The human *PDGFRA* BAC-110kb without CS#1 still express the human *PDGFRA* in postnatal OLPs. In addition, the 2.2kb LacZ human *PDGFRA* plasmid transgenics did not express LacZ in postnatal OLPs (data not shown). These results suggest that CS#1 is not a pOLP-SRE of the human *PDGFRA* gene.

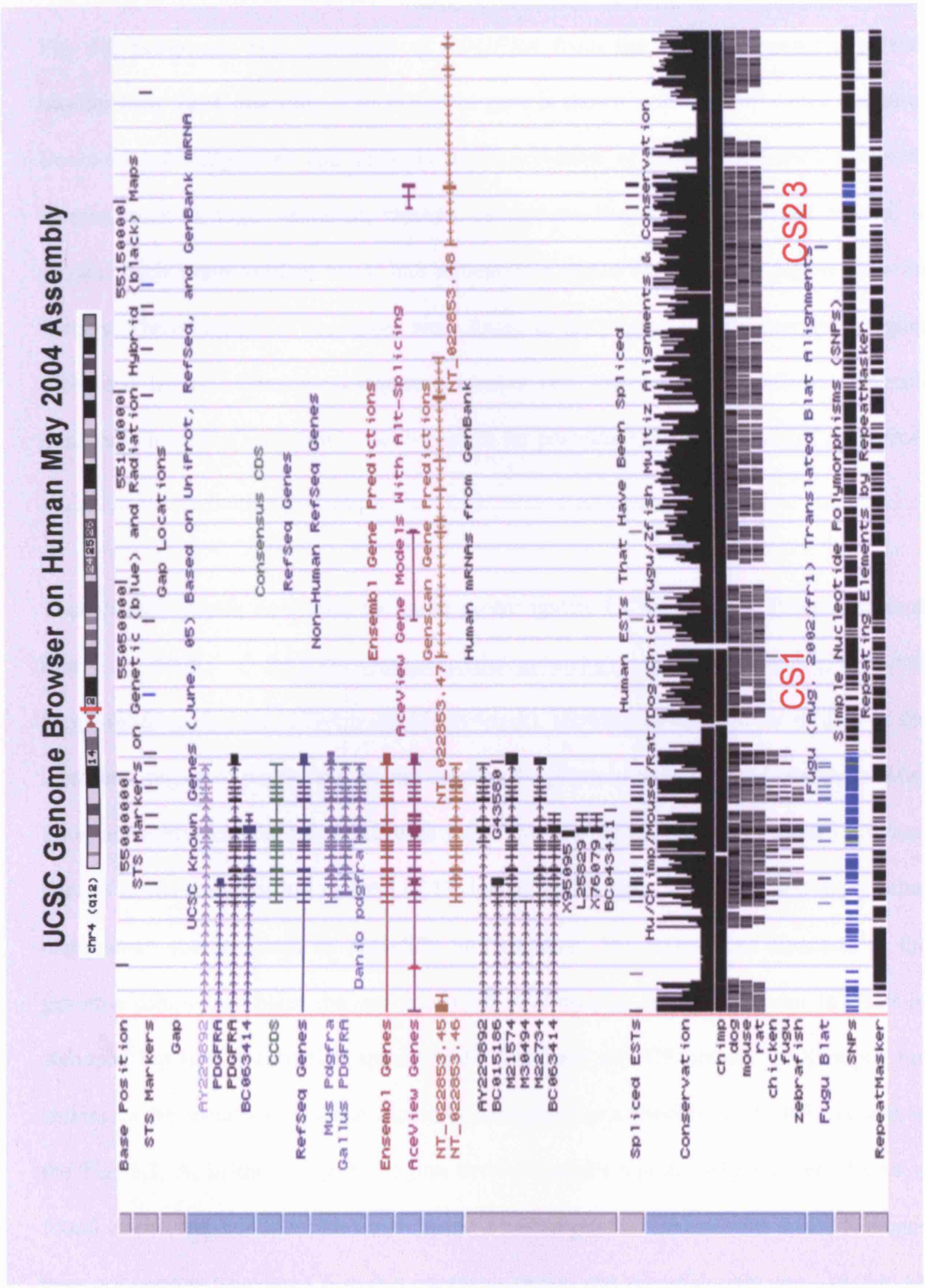

UCSC Genome Browser on Human May 2004 Assembly
chr4 (q12)
CS1
CS23

Fig. 5.2 Schematic representation of *PDGFRA* from the UCSC Genome browser, human May 2004 assembly. The *PDGFRA* gene is shown at the top left of the diagram. Down a multi-alignment and analysis of conservation of sequences between several species, such as fugu, zebrafish, chicken, rat, mouse, dog, chimpanzee and human, is shown. Each small vertical black line represents a conserved sequence region between species. Three conserved sequences were found in the *pdgfra* and *kit* intergenic region (CSs, red letters). Conserved sequence number one was tested in BAC and plasmid transgenic mice but seem not to be important for postnatal OLP expression of *PDGFRA*.

Those results match now with the more recent update UCSC Genome Browser human Mar. 2006 (http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr4:54617686-55307685&hgsid=84052389&multiz17way=pack), showing an updated set of data in the intergenic region of *pdgfra* and *kit* and deleting the previous three CS predicted in the May 2004 assembly (compare Fig. 5.2 with 5.3). The new update also adds more important species to the comparison to a total of 17, including Xenopus, Tetraodon and many other mammalian species, such as armadillo and elephant. The addition of Xenopus to the genomic comparison helps the search of OLP-SRE because *pdgfra* is express in OLPs of Xenopus but not in zebrafish spinal cord. Therefore, any CS present in Xenopus but lacking in the zebrafish sequence should be considered as a possible OLP-SRE. As seen in the Fig. 5.3, A, in the intergenic region between *pdgfra* and *kit* only one set of CSs is found in chicken, ~142.162 kb far from the 3'end of *pdgfra*. On the other hand, Xenopus does not seem to have any CS in this intergenic region and zebrafish only one ~35. 590 kb far from *pdgfra* 3'end. Therefore, if the Xenopus genomic sequence compared with the

other vertebrate sequences is reliable, then our search for OLP-SREs would have to focus on the sequence from intron 14-15 to the 3'end of *PDGFRA* in the future. A close up analysis to this region (Fig. 5.3, B) shows that there are only a few CSs in Xenopus that are not conserved in zebrafish, and those are localized mainly in *pdgfra* exon 23. The addition of CSs through evolution to exon23 sequence from Xenopus to human is shown in this *in silico* analysis (Fig. 5.3, B, red line).

Transgenic experiments analysing specifically the last *PDGFRA* exons from intron 14-15 to exon 23 would show if this region contains any important specific regulatory element, in particular for postnatal OLPs. Detailed study of exon 23 looks now sensible because of the interesting increase in CSs showed in all the vertebrate species from Xenopus to humans, keeping in mind that the comparison with the Xenopus and zebrafish genomic sequence is reliable in this Mar 2006 assembly by the UCSC Genome browser. If it is confirmed that these CS are SREs of the *pdgfra* gene, a TFBS analysis will then reveal more about the TFs involved in its specific transcriptional regulation.


A
UCSC Genome Browser on Human Mar. 2006 Assembly
CS
CS

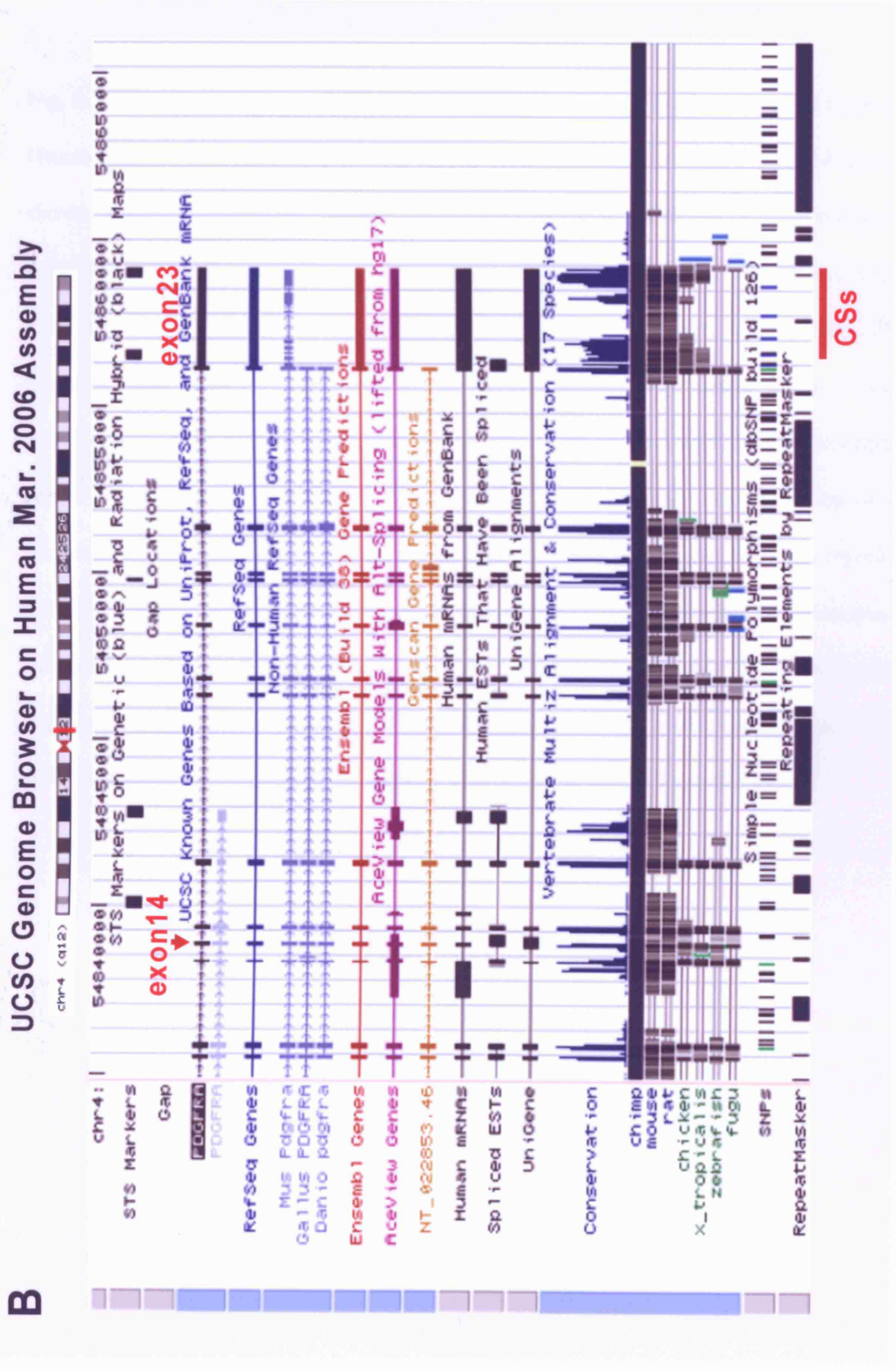

B
UCSC Genome Browser on Human Mar. 2006 Assembly
exon14
exon23
CSs
chr4:
STS Markers
Gap
PDGFRA
RefSeq Genes
Mus Pdgfra
Gallus PDGFRA
Danio pdgfra
Ensembl Genes
AceView Genes
NT_022853.46
Human mRNAs
Spliced ESTs
UniGene
Conservation
chimp
mouse
rat
chicken
X_tropicalis
zebrafish
fugu
SNPs
RepeatMasker

Fig. 5.3. Schematic representations of the *pdgfra* and *kit* genomic sequences from the Human Mar. 2006 assembly of the UCSC Genome browser. The *PDGFRA* gene is shown at the top left of the diagram. A multi-alignment and analysis of conservation of sequences between several species, such as fugu, Tetraodon, Xenopus, zebrafish, chicken, opossum, elephant, rat, mouse, dog, chimpanzee and human, is shown at the bottom of the diagram. Each small vertical black line represents a conserved sequence region between species. **A.** Two conserved sequences (CS) were found in the *pdgfra* and *kit* intergenic region however none is relevant for the search of the OLP-SREs. **B.** A close up of the sequence between intron 14-15 and exon23 sequence shows more detail. In this region of the *pdgfra* gene, the most interesting part showing conserved sequences from Xenopus to human is exon23, which includes the 3'UTR. It seems that through evolution sequences have been added to exon23 consecutively and they have been conserved (CSs, red line), which makes it a very interesting region to analyse for pOLP-SREs in the future.

5.3 Discussion

5.3.1. Prediction of conserved sequences of the *pdgfra* gene during vertebrate evolution

An important part of the genetic basis for the evolution of development are changes in transcriptional regulation (Wray, 2003). Therefore, finding conserve sequences between several species seemed an alternative way to find SREs of important genes during development, such as *pdgfra*.

At the beginning of this PhD research project, it was not possible to make any bioinformatic analysis of the *pdgfra* non-coding sequence, on the scale that is required. First, because the non-coding sequence was not fully sequenced yet for any vertebrate, not even human, and second because the software to analyse such long regions of genomic DNA had not been developed yet. Recently, even though the bioinformatic analysis still provides many false-positive predictions, mainly due to remaining uncompleted genome sequencing projects, enourmous amount of data to process, and arguments about the best mathematical algorithms to use. It is still a valuable starting point for the analysis of the transcriptional regulation of a complex gene. In addition, bioinformatic tools will improve rapidly in the future, although their predictions will always require to be confirmed in biological experimental systems.

In this chapter, I analyse the non-coding sequence upstream, downstream and in the *pdgfra* gene of several vertebrates with bioinformatic tools in other to find conserved sequences that could be *cis*-specific regulatory elements. The first *in silico* search was

done between 100kb non-coding upstream and downstream human and mouse genomic sequences, starting with BLAST multi-allignment and the MATCH tool using the TRANSFAC database to find TFBS in the sequences. Even though the high homology between mouse and human genomic sequence in this analysis that could give many false positive predictions, the finding of FoxJ2 and Pax4 TFBS as possible regulators could be important for *pdgfra* transcription during CNS development. On the other hand, FoxD3 could be important to activate the expression of *pdgfra* in neural crest cells (Dottori et al., 2001). More research about the binding of TFs and possible activation of *pdgfra* gene expression would be important to understand the transcriptional networks established in each cell type to regulate *pdgfra* gene expression during development.

During three years, I tested; 1) other multi-alignment tools that could arrange hundreds of kilobases long non-coding genomic sequences properly; 2) TF databases and tools to find TFBS in the DNA reliably and not producing hundreds of false positive hits; 3) and any bioinformatic genomic browser that could integrate alignments of genomic sequences of many vertebrate species. However, none of them seem to give reliable or precise data to predict any *pdgfra* conserve element that could be interesting to analyse further. Finally, in 2004 the UCSC genome browser included the multialigment and genomic sequence comparison between several species showing a prediction of the conservation of sequences.

Analysing the coding *pdgfra* gene sequence between several species in the UCSC genome browser, it shows that there is high conservation from fugu to human in all exons that encode the tyrosine kinase domains, from exon 12 to exon 23 (Fig 5.3). Similarly,

aminoacid sequence comparison between zebrafish, mouse and human *PDGFRA* showed all tyrosine kinase-domains of this receptor protein are from 100(100) to 70(81) similar (percent identity (similarity)) (Liu et al., 2002). However, the extracellular domain has less similarity between species, from 40(58) in mouse 38(59) in human and 39(61) in Xenopus (Liu et al., 2002). Comparatively, the first exons of the *pdgfra* gene that codify for the extracellular domain are not as well conserved between species, especially exon 1 and exon2. Xenopus, fugu and Tetradon *pdgfra* do not seem to have exon1 represented in the UCSC Genome browser (May. 2006 assembly). It would be interesting to investigate the relevance of this in the pdgfra structure or receptor activity during evolution.

Overall, there is around 74 similarity among human, mouse and Xenopus pdgfra amino acid sequences (Liu et al., 2002). A study in fugu, mainly about *pdgfrb* and its genomic cluster but also about the *pdgfra* gene, shows a phylogenetic tree of vertebrate pdgfra (Williams et al., 2002).

On the other hand, OLPs in the ventral neuroepithelium start expressing *pdgfra* at E12.5 in mice, E14 in rat, E6 in chick and stage 40 in Xenopus. All these vertebrates continue expressing *pdgfra* in OLPs during development and in adulthood. However, zebrafish OLPs never seem to express *pdgfra* (Liu et al., 2002;Park et al., 2002;Mora, 2005). This suggests that a change in the genomic *pdgfra* sequence occurred during the fish-tetrapod transition around 400 million years ago, which activated *pdgfra* expression in OLPs and was conserved subsequently. This could have been an addition of a regulatory element or a mutation to the sequence to activate *pdgfra* transcription specifically in OLPs. Therefore, during my bioinformatic analysis of the *pdgfra* non-coding sequence, I searched

particularly for CSs that seem to be in Xenopus and other vertebrates but not in zebrafish. However, although a downstream CS that fulfilled this requirement was predicted in the May 2004 genome assembly, it was removed later in the Mar. 2006 assembly, showing how difficult it still is to detect functional CSs by this kind of sequence comparison. In fact, there are no CSs predicted in the downstream sequence of *pdgfra* that seem relevant for its transcriptional regulation. Therefore, the search for pOLP-SRE might be best focused in the region from intron 14-15 to exon 23 where more conservation and integration of sequences from fugu to human *pdgfra* seem to have occurred.

5.3.2. Conserved sequences in the 3'UTR of the *pdgfra* gene.

In the last Mar 2006 assembly, there are no CSs predicted in the intergenic region between *pdgfra* and *c-kit* that appeared in Xenopus and were then conserved, this is very surprising considering that it is a sequence region 359.7 kb long. Detailed observation allowed me to see the strong conservation of all *pdgfra* exon sequences except for exon 1 in Xenopus, but very few conserved regions in all *pdgfra* introns. However, it seems that *pdgfra* exon23 sequence has had consecutive addition and conservation of sequences through evolution, from Xenopus to human. These CSs make up to a final ~3.5kb long exon23 sequence in the human *PDGFRA* gene (CSs in Fig 5.3, B). This observation makes it a very interesting region to analyse for SREs of the *pdgfra* gene, and specifically for the pOLP-SREs. In future, more transgenic mice containing up to exon23, or even fragments of it, will be very interesting to analyse in order to continue searching the *pdgfra* SREs.

5.3.3. General conclusion and future experiments.

In conclusion, my extensive analysis of at least 16 different human *PDGFRA* mice transgenic lines during development helped to test *in vivo* the possible locations of SREs in the genomic sequence, and narrow down the search for the pOLP-SREs to a 63 kb region, eliminating 312 kb sequences from the search. In addition, after this *in silico* comparative sequence analysis of the *pdgfra* gene during vertebrate evolution, the most interesting region to keep searching for pOLP-SREs is exon 23, which also contains the 3'UTR, reducing then the search to a ~3.5 kb sequence. Such a short region could be analysed easily and faster experimentally, one by deleting this exon23 from the human *PDGFRA* BAC-110kb sequence (Fig.4.2) by BAC homologous recombination (Chapter 2.6) and adding a reporter gene, or alternatively by fragmenting the same human *PDGFRA* BAC-110kb after exon 23 and detect whether without any downstream sequence the postnatal OLPs keep expressing or not the human *PDGFRA* gene in transgenic mice. *In vitro* experiments with stem cells or 04 cells could later be considered to analyse faster this 3.5 kb *pdgfra* exon23 sequence.

Reference List

1. Aase,K., Abramsson,A., Karlsson,L., Betsholtz,C., and Eriksson,U. (2002). Expression analysis of PDGF-C in adult and developing mouse tissues. Mechanisms of Development *110*, 187-191.

2. Afink,G.J., Nister,M., Stassen,B.H.G.J., Joosten,P.H.L.J., Rademakers,P.J.H., Bongcam-Rudloff,E., Van Zoelen,E.J.J., and Mosselman,S. (1995). Molecular cloning and functional characterization of the human platelet-derived growth factor α receptor gene promoter. Oncogene *10*, 1667-1672.

3. Altaba,A. (1999a). Gli proteins encode context-dependent positive and negative functions: implications for development and disease. Development *126*, 3205-3216.

4. Altaba,A. (1999b). The works of GLI and the power of hedgehog. Nat.Cell Biol. *1*, E147-E148.

5. Altaba,A., Sanchez,P., and Dahmane,N. (2002). Gli and hedgehog in cancer: tumours, embryos and stem cells. Nat.Rev.Cancer *2*, 361-372.

6. Altaba,A., Stecca,B., and Sanchez,P. (2004). Hedgehog--Gli signaling in brain tumors: stem cells and paradevelopmental programs in cancer. Cancer Lett. *204*, 145-157.

7. Arnosti,D.N. (2002). Design and function of transcriptional switches in *Drosophila*. Insect Biochemestry and Molecular Biology *32*, 1257-1273.

8. Battegay,E.J., Raines,E.W., Colbert,T., and Ross,R. (1995). TNF-alpha stimulation of fibroblast proliferation. Dependence on platelet-derived growth factor (PDGF) secretion and alteration of PDGF receptor expression. J.Immunol. *154*, 6040-6047.

9. Baxter,E.J., Hochhaus,A., Bolufer,P., Reiter,A., Fernandez,J.M., Senent,L., Cervera,J., Moscardo,F., Sanz,M.A., and Cross,N.C.P. (2002). The t(4;22)(q12;q11) in atypical chronic myeloid leukaemia fuses BCR to PDGFRA. Human Molecular Genetics *11*, 1391-1397.

10. Berrozpe,G., Agosti,V., Tucker,C., Blanpain,C., Manova,K., and Besmer,P. (2006). A distant upstream locus control region is critical for expression of the Kit receptor gene in mast cells. Mol.Cell Biol. *26*, 5850-5860.

11. Berrozpe,G., Timokhina,I., Yukl,S., Tajima,Y., Ono,M., Zelenetz,A.D., and Besmer,P. (1999). The W(sh), W(57), and Ph Kit expression mutations define tissue-specific control elements located between -23 and -154 kb upstream of Kit. Blood *94*, 2658-2666.

12. Betsholtz,C., Karlsson,L., and Lindahl,P. (2001). Developmental roles of platelet-derived growth factors. BioEssays *23*, 494-507.

13. Bonner,J.C., Badgett,A., Lindroos,P.M., and Coin,P.G. (1996). Basic fibroblast growth factor induces expression of the PDGFreceptor-alpha on human bronchial smooth muscle cells. Am.J.Physiol. *271*, L880-888.

14. Bonner,J.C., Lindroos,P.M., Rice,A.B., Moomaw,C.R., and Morgan,D.L. (1998). Induction of PDGF receptor-alpha in rat myofibroblasts during pulmonary fibrogenesis in vivo. Am.J.Physiol. *274*, L72-80.

15. Bonni,A. and Greenberg,M.E. (1997). Neurotrophin regulation of gene expression. Can.J.Neurol.Sci. *24*, 272-283.

16. Bonni,A., Sun,Y., Nadal-Vicens,M., Bhatt,A., Frank,D.A., Rozovsky,I., Stahl,N., Yancopoulos,G.D., and Greenberg,M.E. (1997). Regulation of gliogenesis in the central nervous system by the JAK-STAT signaling pathway. Science *278*, 477-483.

17. Bostrom,H., Willets,K., Pekny,M., Leveen,P., Lindahl,P., Hedstrand,H., Pekna,M., Hellstrom,M., Gebre-Medhin,S., Schalling,M., and et.al. (1996). PDGF-A signaling is a critical event in lung alveolar myofibroblast development and alveogenesis. Cell *85*, 863-873.

18. Briscoe,J. and Ericson,J. (1999). The specification of neuronal indentity by graded sonic hedgehog signalling. Seminars in Cell and Developmental Biology *10*, 353-362.

19. Briscoe,J., Pierani,A., Jessell,T.M., and Ericson,J. (2000). A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. Cell *101*, 435-445.

20. Bulger,M. and Groudine,M. (1999). Looping versus linking: toward a model for long-distance gene activation. Genes Dev. *13*, 2465-2477.

21. Bulger,M., van Doorninck,J.H., Saitoh,N., Telling,A., Farrell,C., Bender,M.A., Felsenfeld,G., Axel,R., and Groudine,M. (1999). Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc.Natl.Acad.Sci.U.S.A *96*, 5129-5134.

22. Burke,D.T., Carle,G.F., and Olson,M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science *236*, 806-812.

23. Buttitta,L., Mo,R., Hui,C.C., and Fan,C.M. (2003). Interplays of Gli2 and Gli3 and their requirement in mediating Shh-dependent sclerotome induction. Development *130*, 6233-6243.

24. Cai,J., Qi,Y., Hu,X., Tan,M., Liu,Z., Zhang,J., Li,Q., Sander,M., and Qiu,M. (2005). Generation of oligodendrocyte precursor cells from mouse dorsal spinal cord independent of Nkx6 regulation and Shh signaling. Neuron *45*, 41-53.

25. Cameron-Curry,P. and Le Douarin,N.M. (1995). Oligodendrocyte precursors originate from both the dorsal and the ventral parts of the spinal cord. Neuron *15*, 1299-1310.

26. Clement,V., Sanchez,P., de Tribolet,N., Radovanovic,I., and Altaba,A. (2007). HEDGEHOG-GLI1 signaling regulates human glioma growth, cancer stem cell self-renewal, and tumorigenicity. Curr.Biol. *17*, 165-172.

27. Cools,J., De Angelo,D.J., Gotlib,J., Stover,E.H., Legare,R.D., Cortes,J., Kutok,J., Clark,J., Galinsky,I., Griffin,J.D., Cross,N.C.P., Tefferi,A., and others,A.1. (2003). A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. New Eng.J.Med. *348*, 1201-1214.

28. Dahmane,N., Sanchez,P., Gitton,Y., Palma,V., Sun,T., Beyna,M., Weiner,H., and Altaba,A. (2001). The Sonic Hedgehog-Gli pathway regulates dorsal brain growth and tumorigenesis. Development *128*, 5201-5212.

29. Dhar,V., Skoultchi,A.I., and Schildkraut,C.L. (1989). Activation and repression of a β-globin gene in cell hybrids is accompanied by a shift in its temporal replication. Mol.Cell.Biol. *9*, 3524-3532.

30. Di Rocco,F., Carroll,R.S., Zhang,J., and Black,P.M. (1998). Platelet-derived growth factor and its receptor expression in human oligodendrogliomas. Neurosurgery *42*, 341-346.

31. Disteche,C.M., Plowman,G.D., Gronwald,R.G.K., Kelly,J., Bowen-Pope,D., Adler,D.A., and Murray,J.C. (1989). Mapping of the amphiregulin and the platelet-growth factor receptor alpha genes to the proximal long arm of chromosome 4. Cytogenet.Cell Genet. *51*, 990.

32. Dottori,M., Gross,M.K., Labosky,P., and Goulding,M. (2001). The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. Development *128*, 4127-4138.

33. Eccleston,P.A., Funa,K., and Heldin,C.H. (1993). Expression of Platelet-derived growth factor (PDGF) and PDGFα– and β - receptors in the peripheral nervous system: an analysis of sciatic nerve and dorsal root ganglia. Developmental Biology *155*, 459-470.

34. Echelard,Y., Epstein,D.J., St-Jacques,B., Shen,L., Mohler,J., McMahon,J.A., and McMahon,A.P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell *75*, 1417-1430.

35. Ericson,J., Briscoe,J., Rashbass,P., van Heyningen,V., and Jessell,T.M. (1997). Graded sonic hedgehog signaling and the specification of cell fate in the ventral neural tube. Cold Spring Harb Symp Quant Biol *62*, 451-466.

36. Ericson,J., Morton,S., Kawakami,A., Roelink,H., and Jessell,T.M. (1996). Two critical periods of sonic hedgehog signaling requiered for the specification of motor neuron identity. Cell *87*, 661-673.

37. Fan,C.M., Porter,J.A., Chiang,C., Chang,D.T., Beachy,P.A., and Tessier-Lavigne,M. (1995). Long-range sclerotome induction by sonic hedgehog: direct role of the amino-terminal cleavage product and modulation by the cyclic AMP signaling pathway. Cell *81*, 457-465.

38. Fan,C.M. and Tessier-Lavigne,M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. Cell *79*, 1175-1186.

39. Fleming,T.P., Saxena,A., Clark,W.C., Robertson,J.T., Oldfield,E.H., Aaronson,S.A., and Ali,I.U. (1992). Amplification and /or overexpression of platelet-derived growth factor receptors and epidermal growth factor receptor in human glial tumors. Cancer Research *52*, 4550-4553.

40. Fogarty,M., Richardson,W.D., and Kessaris,N. (2005). A subset of oligodendrocytes generated from radial glia in the dorsal spinal cord. Development *132*, 1951-1959.

41. Fruttiger,M., Karlsson,L., Hall,A.C., Abramsson,A., Calver,A.R., Bostrom,H., Willets,K., Bertold,C.H., Heath,J.K., Betsholtz,C., and Richardson,W.D. (1999). Detective oligodendrocyte development and severe hypomyelination in PDGF-A knockout mice. Development *126*, 457-467.

42. Gammill,L.S. and Bronner-Fraser,M. (2003). Neural crest specification: migrating into genomics. Nature *4*, 795-805.

43. Gibson,G. and Honeycutt,E. (2002). The evolution of developmental pathways. Current Opinion in Genetics and Development *12*, 695-700.

44. Gilbert,S.F. (2000). Developmental Biology. (Sunderland: Sinauer), pp. 749.

45. Gilbert,S.F. and Raunio,A.M. (1997). Embryology, constructing the organism. (Sunderland: Sinauer), pp. 537.

46. Giraldo,P. and Montoliu,L. (2001). Size matters: use of YACs, BACs and PACs in transgenic animals. Transgenic Res. *10*, 83-103.

47. Gnessi,L., Basciani,S., Mariani,S., Arizzi,M., Spera,G., Wang,C., Bondjers,C., Karlsson,L., and Betsholtz,C. (2000). Leydig cell loss and spermatogenic arrest in platelet-derived growth factor (PDGF)-A- deficient mice. J.Cell.Biol. *149*, 1019-1025.

48. Gronwald,R.G.K., Adler,D.A., Kelly,J.D., Disteche,C.M., and Bowen-Pope,D.F. (1990). The human PDGF receptor alpha-subunit gene maps to chromosome 4 in close proximity to c-kit. Hum.Genet. *85*, 383-385.

49. Grove,E.A., Williams,B.P., Li,D.Q., Hajihosseini,M., Friedrich,A., and Price,J. (1993). Multiple restricted lineages in the embryonic rat cerebral cortex. Development *117*, 553-561.

50. Gruneberg,H. and Truslove,G.M. (1960). Two closely linked genes in the mouse. Genet.Res.Camb. *1*, 69-90.

51. Hajihosseini,M., Tham,T.N., and Dubois-Dalcq,M. (1996). Origin of oligodendrocytes within the human spinal cord. J.Neurosci. *16*, 7981-7994.

52. Hall,A., Giese,N.A., and Richardson,W.D. (1996). Spinal cord oligodendrocytes develop from ventrally derived progenitor cells that express PDGF alpha-receptors. Development *122*, 4085-4094.

53. Hart,I.K., Richardson,W.D., Heldin,C.-H., Westermark,B., and Raff,M.C. (1989). PDGF receptors on cells of the oligodendrocyte-type-2 astrocyte (O-2A) cell lineage. Development *105*, 595-603.

54. Heldin,C.-H. and Westermark,B. (1999). Mechanism of action and in vivo role of platelet-derived growth factor. Physiol.Rev. *79*, 1283-1316.

55. Helwig,U., Imai,K., Schmahl,W., Thomas,B.E., Varnum,D.S., Nadeau,J.H., and Balling,R. (1995). Interaction between undulated and Patch leads to an extreme form of spina bifida in double-mutant mice. Nature Genetics *11*, 60-63.

56. Hermanson,M., Funa,K., Hartman,M., Claesson-Welsh,L., Heldin,C.H., Westermark,B., and Nister,M. (1992). Platelet-derived growth factor and its receptor in human glioma tissue: Expression of messenger RNA and protein suggests the presence of autocrine and paracrine loops. Cancer Research *52*, 3213-3219.

57. Hermanson,M., Funa,K., Koopmann,J., Maintz,D., Waha,A., Westermark,B., Heldin,C.-H., Wiestler,O.D., Louis,D.N., Von Deimling,A., and Nister,M. (1996). Association of loss of heterozygosity on chromosome 17p with high platelet-derived growth factor α receptor expression in human malignant glioma. Cancer Research *56*, 164-171.

58. Hiller,M.A., Lin,T.Y., Wood,C., and Fuller,M.T. (2001). Developmental regulation of transcription by a tissue-specific TAF homolog. Genes Dev. *15*, 1021-1030.

59. Hoch,R.V. and Soriano,P. (2003). Roles of PDGF in animal development. Development *130*, 4769-4784.

60. Hsieh,C.L., Navankasattusas,S., Escobedo,J.A., Williams,L.T., and Francke,U. (1991). Chromosomal localization of the gene for AA-type platelet-derived growth

factor receptor (PDGFRA) in humans and mice. Cytogenet.Cell Genet. *56*, 160-163.

61. Hu,X., Eszterhas,S., Pallazzi,N., Bouhassira,E.E., Fields,J., Tanabe,O., Gerber,S.A., Bulger,M., Engel,J.D., Groudine,M., and Fiering,S. (2006). Transcriptional interference among the murine {beta}-like globin genes. Blood.

62. Huxley,C. (1998). Exploring gene function: use of yeast artificial chromosome transgenesis. Methods *14*, 199-210.

63. Iseki,S., Araga,A., Ohuchi,H., Nohno,T., Yoshioka,H., Hayashi,F., and Noji,S. (1996). Sonic hedgehog is expressed in epithelial cells during development of whisker, hair, and tooth. Biochem.Biophys.Res.Commun. *218*, 688-693.

64. Jagla,K., Bellard,M., and Frasch,M. (2001). A cluster of Drosophila homeobox genes involved in mesoderm differentiation programs. BioEssays *23*, 125-133.

65. Jessell,T.M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nature Review Genetics *1*, 20-29.

66. Joosten,P.H., Toepoel,M., van Oosterhout,D., Afink,G.B., and van Zoelen,E.J. (2002). A regulating element essential for PDGFRA transcription is recognized by neural tube defect-associated PRX homeobox transcription factors. Biochim.Biophys.Acta *1588*, 254-260.

67. Joosten,P.H.L.J., Hol,F.A., van Beersum,S.E.C., Peters,H., Hamel,B.C.J., Afink,G.J., Van Zoelen,E.J.J., and Mariman,E.C.M. (1998). Altered regulation of platelet-derived growth factor receptor-α gene-transcription *in vitro* by spina bifida-associated mutant Pax1 proteins. Proceedings of the National Academy of Sciences, USA *95*, 14459-14463.

68. Joosten,P.H.L.J., Toepoel,M., Mariman,E.C.M., and Van Zoelen,E.J.J. (2001). Promoter haplotype combinations of the platelet-derived growth factor α-receptor gene predispose to human neural tube defects. Nature Genetics *27*, 215-217.

69. Kalthoff,K. (2001). Analysis of Biological development. (New York: McGraw Hill), pp. 790.

70. Karlsson,L., Bondjers,C., and Betsholtz,C. (1999). Roles for PDGF-A and sonic hedgehog in development of mesenchymal components of the hair follicle. Development *126*, 2611-2621.

71. Karlsson,L., Lindahl,P., Heath,J.K., and Betsholtz,C. (2000). Abnormal gastrointestinal development in PDGF-A and PDGFR-(alpha) deficient mice implicates a novel mesenchymal structure with putative instructive properties in villus morphogenesis. Development *127*, 3457-3466.

72. Kataoka,H., Takakura,N., Nishikawa,S., Tsuchida,K., Kodama,H., Kunisada,T., Risau,W., Kita,T., and Nishikawa,S.I. (1997). Expressions of PDGF receptor

alpha, c-Kit and Flk1 genes clustering in mouse chromosome 5 define distinct subsets of nascent mesodermal cells. Dev.Growth Differ. *39*, 729-740.

73. Kaufman,M.H. and Bard,J.B.L. (1999). The anatomical basis of mouse development. (London: Academic Press), pp. 291.

74. Kawagishi,J., Kumabe,T., Yoshimoto,T., and Yamamoto,T. (1995). Structure, organization, and transcription units of the human α-platelet-derived growth factor receptor gene, PDGFRA. Genomics *30*, 224-232.

75. Kessaris,N., Fogarty,M., Iannarelli,P., Grist,M., Wegner,M., and Richardson,W.D. (2006). Competing waves of oligodendrocytes in the forebrain and postnatal elimination of an embryonic lineage. Nat.Neurosci. *9*, 173-179.

76. Kessaris,N., Pringle,N., and Richardson,W.D. (2001). Ventral neurogenesis and the neuron-glial switch. Neuron *31*, 677-680.

77. Kitami,Y., Inui,H., Uno,S., and Inagami,T. (1995). Molecular structure and transcriptional regulation of the gene for the platelet-derived growth factor alpha receptor in cultured vascular smooth muscle cells. J.Clin.Invest. *96*, 558-567.

78. Klinghoffer,R.A., Mueting-Nelsen,P.F., Faerman,A., Shani,M., and Soriano,P. (2001). The two PDGF receptors maintain conserved signaling in vivo despite divergent embryological functions. Molec.Cell *7*, 343-354.

79. Kluppel,M., Nagle,D.L., Bucan,M., and Bernstein,A. (1997). Long-range genomic rearrangements upstream of Kit dysregulate the developmental pattern of Kit expression in W57 and Wbanded mice and interfere with distinct steps in melanocyte development. Development *124*, 65-77.

80. Kraft,H.J., Mosselman,S., Smits,H.A., Hohenstein,P., Piek,E., Chen,Q., Artzt,K., and van Zoelen,E.J. (1996). Oct-4 regulates alternative platelet-derived growth factor alpha receptor gene promoter in human embryonal carcinoma cells. J.Biol.Chem. *271*, 12873-12878.

81. Kumabe,T., Sohma,Y., Kayama,T., Yoshimoto,T., and Yamamoto,T. (1992). Amplification of α–platelet-derived growth factor receptor gene lacking an exon coding for a portion of the extracellular region in a primary brain tumor of glial origin. Oncogene *7*, 627-633.

82. Lee,E.C., Yu,D., Martinez,d., V, Tessarollo,L., Swing,D.A., Court DL, Jenkins,N.A., and Copeland,N.G. (2001). A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics *73*, 56-65.

83. Leveen,P., Pekny,M., Gebre-Medhin,S., Swolin,B., Larsson,E., and Betsholtz,C. (1994). Mice deficient for PDGF B show renal, cardiovascular, and hematological abnormalities. Genes and Development *8*, 1875-1887.

84. Levison,S.W., Chuang,C., Abramson,B.J., and Goldman,J.E. (1993). The migrational patterns and developmental fates of glial precursors in the rat subventricular zone are temporally regulated. Development *119*, 611-622.

85. Levison,S.W. and Goldman,J.E. (1993). Both oligodendrocytes and astrocytes develop from progenitors in the subventricular zone of postnatal rat forebrain. Neuron *10*, 201-212.

86. Li,L., Schatteman,G.C., Oppenheim,R.W., Lei,M., Bowen-Pope,D.F., and Houenou,L.J. (1996). Altered development of spinal cord in the mouse mutant (Patch) lacking the PDGF receptor α - subunit gene. Developmental Brain Research *96*, 204-209.

87. Lih,C.-J., Cohen,S.N., Wang,C., and Lin-Chao,S. (1996). The platelet-derived growth factor α-receptor is encoded by a growth-arrest-specific (*gas*) gene. Proceedings of the National Academy of Sciences, USA *93*, 4617-4622.

88. Lindahl,P., Hellstrom,M., Kalen,M., Karlsson,L., Pekny,M., Pekna,M., Soriano,P., and Betsholtz,C. (1998). Paracrine PDGF-B/PDGF-Rbeta signaling controls mesangial cell development in kidney glomeruli. Development *125*, 3313-3322.

89. Lindroos,P.M., Coin,P.G., Badgett,A., Morgan,D.L., and Bonner,J.C. (1997). Alveolar macrophages stimulated with titanium dioxide, chrysotile asbestos, and residual oil fly ash upregulate the PDGF receptor-alpha on lung fibroblasts through an IL-1 beta-dependent mechanism. Am.J.Respir.Cell.Mol.Biol. *16*, 283-292.

90. Liu,L., Chong,S.-W., Balasubramaniyan,N.V., Korzh,V., and Ge,R. (2002). Platelet-derived growth factor receptor alpha (*pdgrf-α*) gene in zebrafish embryonic development. Mechanisms of Development *116*, 227-230.

91. Lu,Q.R., Cai,L., Rowitch,D., Cepko,C.L., and Stiles,C.D. (2001). Ectopic expression of Olig1 promotes oligodendrocyte formation and reduces neuronal survival in developing mouse cortex. Nat.Neurosci. *4*, 973-974.

92. Lu,Q.R., Sun,T., Zhu,Z., Ma,N., Garcia,M., Stiles,C.D., and Rowitch,D.H. (2002). Common developmental requirement for *Olig* function indicates a motor neuron/oligodendrocyte connection. Cell *109*, 75-86.

93. Lu,Q.R., Yuk,D., Alberta,J.A., Zhu,Z., Pawlintzky,I., Chan,J., McMahon,A.P., Stiles,C.D., and Rowitch,D.H. (2000). Sonic hedgehog-regulated oligodendrocyte lineage genes encoding bHLH proteins in the mammalian central nervous system. Neuron *25*, 317-329.

94. Marquardt,T. and Pfaff,S.L. (2001). Cracking the transcriptional code for cell specification in the neural tube. Cell *106*, 651-654.

95. Matsui,T., Heidaran,M., Miki,T., Popescu,N., La Rochelle,W., Kraus,M., Pierce,J., and Aaronson,S. (1989). Isolation of a novel receptor cDNA establishes the existence of two PDGF receptor genes. Science *243*, 800-804.

96. Matys,V., Fricke,E., Geffers,R., Gobling,E., Haubrock,M., Hehl,R., Hornischer,K., Karas,D., Kel,A.E., Kel-Margoulis,O.V., Kloos,D.-U., Land,S., Lewicki-Potapov,B., Michael,H., Munch,R., Reuter,I., Rotert,S., Saxel,H., Scheer,M., Thiele,S., and Wingender,E. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acid Research *31*, 374-378.

97. McKinnon,R.D., Matsui,T., Dubois-Dalcq,M., and Aaronson,S.A. (1990). FGF modulates the PDGF-driven pathway of oligodendrocyte development. Neuron *5*, 603-614.

98. Mercola,M., Melton,D.A., and Stiles,C.D. (1988). Platelet-derived growth factor A chain is maternally encoded in Xenopus embryos. Science *241*, 1223-1225.

99. Mercola,M., Wang,C.Y., Kelly,J., Brownlee,C., Jackson-Grusby,L., Stiles,C., and Bowen-Pope,D. (1990). Selective expression of PDGF A and its receptor during early mouse embryogenesis. Dev.Biol. *138*, 114-122.

100. Monsoro-Burq,A.H. (2005). Sclerotome development and morphogenesis: when experimental embryology meets genetics. Int.J.Dev.Biol. *49*, 301-308.

101. Montoliu,L., Schedl,A., Kelsey,G., Lichter,P., Larin,Z., Lehrach,H., and Schutz,G. (1993). Generation of transgenic mice with yeast artificial chromosomes. Cold Spring Harb.Symp.Quant.Biol. *58*, 55-62.

102. Mora,A. Studies of Gliogenesis in the Central Nervous System of Zebrafish. 1-162. 2005. University College London.
Ref Type: Thesis/Dissertation

103. Morrison-Graham,K., Schatteman,G.C., Bork,T., Bowen-Pope,D.F., and Weston,J.A. (1992). A PDGF receptor mutation in the mouse (Patch) perturbs the development of a non-neuronal subset of neural crest-derived cells. Development *115*, 133-142.

104. Mosselman,S., Claesson-Welsh,L., Kamphuis,J.S., and Van Zoelen,E.J.J. (1994). Developmentally regulated expression of two novel platelet-derived growth factor α-receptor transcripts in human teratocarcinoma cells. Cancer Research *54*, 220-225.

105. Mosselman,S., Looijenga,L.H., Gillis,A.J., van Rooijen,M.A., Kraft,H.J., van Zoelen,E.J., and Oosterhuis,J.W. (1996). Aberrant platelet-derived growth factor alpha-receptor transcript as a diagnostic marker for early human germ cell tumors of the adult testis. Proc.Natl.Acad.Sci.U.S.A *93*, 2884-2888.

106. Nagy,A., Gertsenstein,M., Vintersten,K., and Behringer,R. (2003). Manipulating the Mouse Embryo. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.).

107. Nery,S., Wichterle,H., and Fishell,G. (2001). Sonic hedgehog contributes to oligodendrocyte specification in the mammalian forebrain. Development *128*, 527-540.

108. Newgreen,D.F. and Erickson,C.A. (1986). The migration of neural crest cells. Int.Rev.Cytol. *103*, 89-145.

109. Nister,M., Claesson-Welsh,L., Eriksson,A., Heldin,C.-H., and Westermark,B. (1991). Differential expression of platelet-derived growth factor receptors in human malignant glioma cell lines. J.Biol.Chem. *266*, 16755-16763.

110. Noll,E. and Miller,R.H. (1993). Oligodendrocyte precursors originate at the ventral ventricular zone dorsal to the ventral midline region in the embryonic rat spinal cord. Development *118*, 563-573.

111. Oppenheim,R.W., Prevette,D., Haverkamp,L.J., Houenou,L.J., Yin,Q.-W., and McManaman,J. (1993). Biological studies of a putative avian muscle-derived neurotrophic factor that prevents naturally occurring motoneuron death *in vivo*. Journal of Neurobiology *24*, 1065-1079.

112. Orentas,D.M., Hayes,J.E., Dyer,K.L., and and Miller,R.H. (1999). Sonic hedgehog signaling is required during the appearance of spinal cord oligodendrocyte precursors. Development *126*, 2419-2429.

113. Orr-Urtreger,A., Bedford,M.T., Do,M.S., Eisenbach,L., and Lonai,P. (1992). Developmental expression of the alpha receptor for platelet-derived growth factor, which is deleted in the embryonic lethal Patch mutation. Development *115*, 289-303.

114. Orr-Urtreger,A. and Lonai,P. (1992). Platelet-derived growth factor-A and its receptor are expressed in separate, but adjacent cell layers of the mouse embryo. Development *115*, 1045-1058.

115. Palmieri,S.L., Payne,J., Stiles,C.D., Biggers,J.D., and Mercola,M. (1992). Expression of mouse PDGF-A and PDGF alpha-receptor genes during pre- and post-implantation development: evidence for a developmental shift from an autocrine to a paracrine mode of action. Mech.Dev. *39*, 181-191.

116. Park,H.-C., Mehta,A., Richardson,J.S., and Appel,B. (2002). olig2 is required for zebrafish primary motor neuron and oligodendrocyte development. Developmental Biology *248*, 356-368.

117. Park,H.L., Bai,C., Platt,K.A., Matise,M.P., Beeghly,A., Hui,C.c., Nakashima,M., and Joyner,A.L. (2000). Mouse *Gli1* mutants are viable but have defects in SHH signaling in combination with *Gli2* mutation. Development *127*, 1593-1605.

118. Perris,R., Paulsson,M., and Bronner-Fraser,M. (1989). Molecular mechanisms of avian neural crest cell migration on fibronectin and laminin. Dev.Biol. *136*, 222-238.

119. Pirrotta,V. (1998). Polycombing the genome: PcG, trxG, and chromatin silencing. Cell *93*, 333-336.

120. Pringle,N., Guthrie,S., Lumsden,A., and Richardson,W.D. (1998). Dorsal spinal cord neuroepithelium generates astrocytes but not oligodendrocytes. Neuron *20*, 883-893.

121. Pringle,N.P., Mudhar,H.S., Collarini,E.J., and Richardson,W.D. (1992). PDGF receptors in the rat CNS: during late neurogenesis, PDGF alpha-receptor expression appears to be restricted to glial cells of the oligodendrocyte lineage. Development *115*, 535-551.

122. Pringle,N.P. and Richardson,W.D. (1993). A singularity of PDGF alpha-receptor expression in the dorsoventral axis of the neural tube may define the origin of the oligodendrocyte lineage. Development *117*, 525-533.

123. Pringle,N.P., Yu,W.P., Howell,M., Colvin,J.S., Ornitz,D.M., and Richardson,W.D. (2003). Fgfr3 expression by astrocytes and their precursors: evidence that astrocytes and oligodendrocytes originate in distinct neuroepithelial domains. Development *130*, 93-102.

124. Pringle,N.P., Yu,W.-P., Guthrie,S., Roelink,H., Lumsden,A., Peterson,A.C., and Richardson,W.D. (1996). Determination of Neuroepithelial cell fate: induction of the Oligodendrocyte lineage by ventral midline cells and sonic hedgehog. Developmental Biology *177*, 30-42.

125. Pritsker,M., Liu,Y.-C., Beer,M.A., and Tavazoie,S. (2004). Whole-genome discovery of transcription factor binding sites by network-level conservation. Genome Research *14*, 99-108.

126. Qi,Y., Cai,J., Wu,Y., Wu,R., Lee,J., Fu,H., Rao,M., Sussel,L., Rubenstein,J., and Qiu,M. (2001). Control of oligodendrocyte differentiation by the Nkx2.2 homeodomain transcription factor. Development *128*, 2723-2733.

127. Qi,Y., Tan,M., Hui,C.-C., and Qiu,M. (2003). *Gli2* is required for normal Shh signaling and oligodendrocyte development in the spinal cord. Molecular and Cellular Neurosciences *23*, 440-450.

128. Raff,M.C. (1989). Glial cell diversification in the rat optic nerve. Science *243*, 1450-1455.

129. Raff,M.C., Miller,R.H., and Noble,M. (1983). A glial progenitor cell that develops in vitro into an astrocyte or an oligodendrocyte depending on culture medium. Nature *303*, 390-396.

130. Ragoczy,T., Bender,M.A., Telling,A., Byron,R., and Groudine,M. (2006). The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. Genes Dev. *20*, 1447-1457.

131. Rappolee,D.A., Brenner,C.A., Schultz,R., Mark,D., and Werb,Z. (1988). Developmental expression of PDGF, TGF-alpha, and TGF-beta genes in preimplantation mouse embryos. Science *241*, 1823-1825.

132. Reinertsen,K.K., Bronson,R.T., Stiles,C.D., and Wang,C. (1997). Temporal and spatial specificity of PDGF alpha receptor promoter in transgenic mice. Gene Expr. *6*, 301-314.

133. Richardson,W.D. (2001). Oligodendrocyte development. In Glial cell development, K. R. Jessen and W. D. Richardson, eds. (Oxford, NY: Oxford University Press), pp. 21-54.

134. Richardson,W.D., Kessaris,N., and Pringle,N. (2006). Oligodendrocyte wars. Nat.Rev.Neurosci. *7*, 11-18.

135. Richardson,W.D., Pringle,N., Yu,W.-P., and Hall,A.C. (1997). Origins of spinal cord oligodendrocytes: possible developmental and evolutionary relationships with motor neurons. Developmental Neurosciences *19*, 58-68.

136. Richardson,W.D., Smith,H.K., Sun,T., Pringle,N., Hall,A., and Woodruff,R. (2000). Oligodendrocyte lineage and motor neuron connection. Glia *29*, 136-142.

137. Roberts,W.M., Look,A.T., Ruossel,M.F., and Sherr,C.J. (1988). Tandem linkage of human CSF-1 receptor (c-fms) and PDGF receptor genes. Cell *55*, 655-661.

138. Roelink,H. (1996). Tripartite signaling of pattern: interactions between Hedgehogs, BMPs and Wnts in the control of vertebrate development. Curr.Opin.Neurobiol. *6*, 33-40.

139. Roelink,H., Augsburger,A., Heemskerk,J., Korzh,V., Norlin,S., Altaba,A., Tanabe,Y., Placzek,M., Edlund,T., Jessell,T.M., and . (1994). Floor plate and motor neuron induction by vhh-1, a vertebrate homolog of hedgehog expressed by the notochord. Cell *76*, 761-775.

140. Roelink,H., Porter,J.A., and et al. (1995). Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. Cell *81*, 445-455.

141. Ross,R., Raines,E.W., and Bowen-Pope,D.F. (1986). The biology of platelet-derived growth factor. Cell *46*, 155-169.

142. Rowitch,D.H., Lu,Q.R., Kessaris,N., and Richardson,W.D. (2002). An 'oligarchy' rules neural development. TRENDS in Neurosciences *25*, 417-422.

143. Samanta,J. and Kessler,J.A. (2004). Interactions between ID and OLIG proteins mediate the inhibitory effects of BMP4 on oligodendroglial differentiation. Development *131*, 4131-4142.

144. Sambrook,J., Fritsch,E.F., and Maniatis,T. (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press).

145. Sanchez,P. and Altaba,A. (2005). In vivo inhibition of endogenous brain tumors through systemic interference of Hedgehog signaling in mice. Mech.Dev. *122*, 223-230.

146. Sanchez,P., Hernandez,A.M., Stecca,B., Kahler,A.J., DeGueme,A.M., Barrett,A., Beyna,M., Datta,M.W., Datta,S., and Altaba,A. (2004). Inhibition of prostate cancer proliferation by interference with SONIC HEDGEHOG-GLI1 signaling. Proc.Natl.Acad.Sci.U.S.A *101*, 12561-12566.

147. Santini,S., Boore,J.L., and Meyer,A. (2003). Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. Genome Res. *13*, 1111-1122.

148. Sasaki,H. and Hogan,B.L.M. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. Development *118*, 47-59.

149. Schatteman,G.C., Morrison-Graham,K., van Koppen,A., Weston,J.A., and Bowen-Pope,D.F. (1992). Regulation and role of PDGF receptor α-subunit expression during embryogenesis. Development *115*, 123-131.

150. Schedl,A., Grimes,B., and Montoliu,L. (1996). YAC Transfer by Microinjection. In YAC PROTOCOLS, D. Markie, ed. (Totowa, New Jersey: Human Press), pp. 378.

151. Schedl,A., Larin,Z., Montoliu,L., Thies,E., Kelsey,G., Lehrach,H., and Schutz,G. (1993a). A method for the generation of YAC transgenic mice by pronuclear microinjection. Nucleic Acids Res. *21*, 4783-4787.

152. Schedl,A., Montoliu,L., Kelsey,G., and Schutz,G. (1993b). A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic mice. Nature *362*, 258-261.

153. Sham,M.H., Vesque,C., Nonchev,S., Marshall,H., Frain,M., Das Gupta,R., Whiting G.J., Wilkinson,D., Charnay,P., and Krumlauf,R. (1993). The zinc finger gene Krox-20 regulates Hox B2 (Hox 2.8) during hindbrain segmentation. Cell *72*, 183-196.

154. Shoshan,Y., Nishiyama,A., Chang,A., Mork,S., Barnett,G.H., Cowell,J.K., Trapp,B.D., and Staugaitis,S.M. (1999). Expression of oligodendrocyte progenitor cell antigens by gliomas: implications for the histogenesis of brain tumors. Proceedings of the National Academy of Sciences, USA *96*, 10361-10366.

155. Silverman G A (1996). End-Rescue of YAC Clone Inserts by Inverse PCR. In YAC PROTOCOLS, D. Markie, ed. (Totowa, New Jersey: Human Press), pp. 378.

156. Sippel,A., Saueressig,H., Winter,D., Grewal,T., Faust,N., Hecht,A., and Bonifer,C. The regulatory domain organization of eukaryotic genomes: implications for stable gene transfer. Grosveld, F. and Kolias, G. 1992. London, Academic Press.
Ref Type: Serial (Book,Monograph)

157. Smith,E.A., Seldin,M.F., Martinez,L., Watson,M.L., Ghosh Choudhury,G., Lalley,P.A., Pierce,J., Aaronson,S., Barker,J., Naylor,S.L., and Sakaguchi,A.Y. (1991). Mouse platelet-derived growth factor receptor alpha gene is deleted in W-

19H and patch mutations on chromosome 5. Proceedings of the National Academy of Sciences, USA 4811-4815.

158. Soo-Kyung,L. and Pfaff,S.L. (2001). Transcriptional netrworks regulating neuronal identity in the developing spinal cord. Nature neuroscience supplement *4*, 1183-1191.

159. Soriano,P. (1994). Abnormal kidney development and hematological disorders in PDGF beta-receptor mutant mice. Genes Dev. *8*, 1888-1896.

160. Soriano,P. (1997). The PDGF alpha receptor is required for neural crest cell development and for normal patterning of the somites. Development *124*, 2691-2700.

161. Soriano,P., Friedrich,G., and Lawinger,P. (1991). Promoter interactions in retrovirus vectors introduced into fibroblasts and embryonic stem cells. J.Virol. *65*, 2314-2319.

162. Soula,C., Danesin,C., and et al. (2001). Distinct sites of origin of oligodendrocytes and somatic motoneurons in the chick spinal cord: oligodendrocytes arise from Nkx2.2-expressing progenitors by a Shh-dependent mechanism. Development *128*, 1369-1379.

163. Spassky,N., Goujet-Zalc,C., Parmantier,E., Olivier,C., Martinez,S., Ivanova,A., Ikenaka,K., Macklin,W., Cerruti,I., Zalc,B., and Thomas,J.-L. (1998). Multiple restricted origin of oligodendrocytes. The Journal of Neuroscience *18*, 8331-8343.

164. Spassky,N., Heydon,K., Mangatan,A., Jankovski,A., Olivier,C., Queraud-Lesaux,F., Goujet-Zalc,C., Thomas,J.L., and Zalc,B. (2001). Sonic hedgehog-dependent emergence of oligodendrocytes in the telencephalon: evidence for a source of oligodendrocytes in the olfactory bulb that is independent of PDGFRα signaling. Development *128*, 4993-5004.

165. Spassky,N., Olivier,C., Perez-Villegas,E., Goujet-Zalc,C., Martinez,S., Thomas,J.-L., and Zalc,B. (2000). Single or multiple oligodendroglial lineages: A controversy. Glia *29*, 143-148.

166. Spitz,F. and Duboule,D. (2005). Developmental biology: reproduction in clusters. Nature *434*, 715-716.

167. Spitz,F., Gonzalez,F., and Duboule,D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell *113*, 405-417.

168. Spitz,F., Gonzalez,F., Peichel,C., Vogt,T.F., Duboule,D., and Zakany,J. (2001). Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. Genes Dev. *15*, 2209-2214.

169. Stecca,B. and Altaba,A. (2005). Brain as a paradigm of organ growth: Hedgehog-Gli signaling in neural stem cells and brain tumors. J.Neurobiol. *64*, 476-490.

170. Stecca,B., Mas,C., and Altaba,A. (2005). Interference with HH-GLI signaling inhibits prostate cancer. Trends Mol.Med. *11*, 199-203.

171. Stephenson,D.A., Mercola,M., Anderson,E., Wang,C., Stiles,C.D., Bowen-Pope,D.F., and Chapman,V.M. (1991). Platelet-derived growth factor receptor alpha-subunit gene (Pdgfra) is deleted in the mouse patch (Ph) mutation. Proceedings of the National Academy of Sciences, USA *88*, 6-10.

172. Strawn,L.M., Mann,E., Elliger,S.S., Chu,L.M., Germain,L.L., Niederfellner,G., Ullrich,A., and Shawver,L.K. (1994). Inhibition of glioma cell growth by a truncated platelet-derived growth factor β receptor. J.Biol.Chem. *269*, 21215-21222.

173. Sun,T. Function and regulation of Platelet-derived growth factor receptor alpha during development. 1-192. 1999. University College London.
Ref Type: Thesis/Dissertation

174. Sun,T., Dong,H., Wu,L., Kane,M., Rowitch,D.H., and Stiles,C.D. (2003). Cross-repressive interaction of the Olig2 and Nkx2.2 transcription factors in developing neural tube associated with formation of a specific physical complex. The Journal of Neuroscience *23*, 9547-9556.

175. Sun,T., Echelard,Y., Lu,R., Yuk,D., Kaing,S., Stiles,C.D., and Rowitch,D.H. (2001). Olig bHLH proteins interact with homeodomain proteins to regulate cell fate acquisition in progenitors of the ventral neural tube. Current Biology *11*, 1413-1420.

176. Sun,T., Jayatilake,D., Afink,G.J., Ataliotis,P., Nister,M., Richardson,W.D., and Smith,H.K. (2000). A human YAC transgene rescues craniofacial and neural tube development in PDGFRα in prenatal lung growth. Development *127*, 4519-4529.

177. Takakura,N., Yoshida,H., Kunisada,T., Nishikawa,S., and Nishikawa,S.I. (1996). Involvement of platelet-derived growth factor receptor-alpha in hair canal formation. J.Invest Dermatol. *107*, 770-777.

178. Takakura,N., Yoshida,H., Ogura,Y., Kataoka,H., Nishikawa,S., and Nishikawa,S. (1997). PDGFR alpha expression during mouse embryogenesis: immunolocalization analyzed by whole-mount immunohistostaining using the monoclonal anti-mouse PDGFR alpha antibody APA5. J.Histochem.Cytochem. *45*, 883-893.

179. Tallquist,M.D. and Soriano,P. (2003). Cell autonomous requirement for PDGFRα in populations of cranial and cardiac neural crest cells. Development *130*, 507-518.

180. Tanabe,Y. and Jessell,T.M. (1996). Diversity and pattern in the developing spinal cord. Science *274*, 1115-1123.

181. Tekki-Kessaris,N., Woodruff,R., Hall,A.C., Gaffield,W., Kimura,S., Stiles,C.D., Rowitch,D.H., and Richardson,W.D. (2001). Hedgehog-dependent oligodendrocyte lineage specification in the telencephalon. Development *128*, 2545-2554.

182. Timsit,S., Martinez,S., Allinquant,B., Peyron,F., Puelles,L., and Zalc,B. (1995). Oligodendrocytes originate in a restricted zone of the embryonic ventral neural tube defined by DM-20 mRNA expression. J.Neurosci. *15*, 1012-1024.

183. Timsit,S.G., Bally-Cuif,L., Colman,D.R., and Zalc,B. (1992). DM-20 mRNA is expressed during the embryonic development of the nervous system of the mouse. J.Neurochem. *58*, 1172-1175.

184. Vallstedt,A., Klos,J.M., and Ericson,J. (2005). Multiple dorsoventral origins of oligodendrocyte generation in the spinal cord and hindbrain. Neuron *45*, 55-67.

185. van Limpt,V. Characterization, isolation and modification of PDGFRA containing YACS for the generation of transgenic mice. 1-58. 1998. Department of Genetics and pathology. Uppsala Universitet, Sweden.
Ref Type: Report

186. Vandenbark,G.R., deCastro,C.M., Taylor,H., Dew-Knight,S., and Kaufman,R.E. (1992). Cloning and structural analysis of the human c-kit gene. Oncogene *7*, 1259-1266.

187. Wang,C. and Stiles,C.D. (1994). Platelet-derived growth factor α receptor gene expression: isolation and characterization of the promoter and upstream regulatory elements. Proceedings of the National Academy of Sciences, USA *91*, 7061-7065.

188. Warf,B.J., Fok-Seang,J., and Miller,R.H. (1991). Evidence for the ventral origin of oligodendrocyte precursors in the rat spinal cord. The Journal of Neuroscience *11*, 2477-2488.

189. Wessels E. Generation and analysis of fragmented yeast artificial chromosomes to characterize oligodendrocyte precursor-specific regulatory elements for platelet-derived growth factor receptor-α expression. Afink, G. B. 1-50. 2001. Department of Genetics and Pathology, Rudbeck Laboratory. University Hospital, University of Upsala, Sweden.
Ref Type: Report

190. Weston,J.A., Yoshida,H., Robinson,V., Nishikawa,S., Fraser,S.T., and Nishikawa,S. (2004). Neural crest and the origin of ectomesenchyme: neural fold heterogeneity suggests an alternative hypothesis. Dev.Dyn. *229*, 118-130.

191. Wilkinson,D.G., Bhatt,S., Chavrier,P., Bravo,R., and Charnay,P. (1989). Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. Nature *337*, 461-464.

192. Williams,H., Brenner,S., and Venkatesh,B. (2002). Identification and analysis of additional copies of the platelet-derived growth factor receptor and colony stimulating factor 1 receptor genes in fugu. Gene *295*, 255-264.

193. Wingender,E., Chen,X., Fricke,E., Geffers,R., Hehl,R., Liebich,I., Krull,M., Matys,V., Michael,H., Ohnhauser,R., Prub,M., Schacherer,F., Thiele,S., and Urbach,S. (2001). The TRANSFAC system on gene expression regulation. Nucleic Acid Research *29*, 281-283.

194. Wolfensohn,S. and Lloyd,M. (1998). Handbook of Laboratory Animal Management and welfare. Blackwell Science Ltd.), pp. -334.

195. Wolpert,L., Beddington,R., Brockes,J., Jessell,T., Lawrence,P., and Meyerowitz,E. (1998). Principles of development. (Oxford: Oxford University Press), pp. 485.

196. Woodruff,R., Tekki-Kessaris,N., Stiles,C.D., Rowitch,D.H., and Richardson,W.D. (2001). Oligodendrocyte development in the spinal cord and telencephalon: common themes and new perspectives. International Journal of Developmental Neuroscience *19*, 379-385.

197. Wray,G.A. (2003). Transcriptional regulation and the evolution of development. International Journal of Developmental Biology *47*, 675-684.

198. Xie,J., Aszterbaum,M., Zhang,X., Bonifas,J.M., Zachary,C., Epstein,E., and McCormick,F. (2001). A role of PDGFRα in basal cell carcinoma proliferation. Proceedings of the National Academy of Sciences, USA *98*, 9255-9259.

199. Yamakage,A., Kikuchi,K., Smith,E.A., Le Roy,E.C., and Trojanowska,M. (1992). Selective upregulation of platelet-derived growth factor α receptors by transforming growth factor β in scleroderma fibroblasts. J.Exp.Med. *175*, 1227-1234.

200. Yeh,H.-J., Ruit,K.G., Wang,Y.-X., Parks,W.C., Snider,W.D., and Deuel,T.F. (1991). PDGF A-chain gene is expressed by mammalian neurons during development and in maturity. Cell *64*, 209-216.

201. Yeh,H.-J., Silos-Santiago,I., Wang,Y.-W., George,R.J., Snider,W.D., and Deuel,T.F. (1993). Developmental expression of the platelet-derived growth factor α-receptor gene in mammalian central nervous system. Proceedings of the National Academy of Sciences, USA *90*, 1952-1956.

202. Yu,D., Ellis,H.M., Lee,E.C., Jenkins,N.A., Copeland,N.G., and Court DL (2000). An efficient recombination system for chromosome engineering in Escherichia coli. Proc.Natl.Acad.Sci.U.S.A *97*, 5978-5983.

203. Yu,W.P., Collarini,E.J., Pringle,N.P., and Richardson,W.D. (1994). Embryonic expression of myelin genes: evidence for a focal source of oligodendrocyte precursors in the ventricular zone of the neural tube. Neuron *12*, 1353-1362.

204. Zardoya,R., Abouheif,E., and Meyer,A. (1996). Evolutionary analyses of hedgehog and *Hoxd-10* genes in fish species closely related to the zebrafish. Proceedings of the National Academy of Sciences, USA *93*, 13036-13041.

205. Zhang,X.Q., Afink,G.B., Svensson,K., Jacobs,J.J., Gunther,T., Forsberg-Nilsson,K., Van Zoelen,E.J.J., Westermark,B., and Nister,M. (1998). Specific expression in mouse mesoderm-and neural crest derived tissues of a human PDGFRa promoter/lacZ transgene. Mechanisms of Development *70*, 167-180.

206. Zhou,Q. and Anderson,D.J. (2002). The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. Cell *109*, 61-73.

207. Zhou,Q., Choi,G., and Anderson,D.J. (2001). The bHLH transcription factor Olig2 promotes oligodendrocyte differentiation in collaboration with Nkx2.2. Neuron *31*, 791-807.

208. Zhou,Q., Wang,S., and Anderson,D.J. (2000). Identification of a novel family of oligodendrocyte lineage-specific basic helix-loop-helix transcription factors. Neuron *25*, 331-343.

209. Zweigerdt,R., Braun,T., and Arnold,H.H. (1997). Faithful expression of the Myf-5 gene during mouse myogenesis requires distant control regions: a transgene approach using yeast artificial chromosomes. Dev.Biol. *192*, 172-180.